Propuesta para múltiples diseños de tiendas basados en reglas de asociación

Andrés Felipe Arboleda Correa

Universidad Nacional de Colombia

Facultad de Minas, Departamento de Ciencias de la Computación y de la Decisión

Medellín, Colombia

2022

Propuesta para múltiples diseños de tiendas basados en reglas de asociación

Andrés Felipe Arboleda Correa

Tesis o trabajo de investigación presentada(o) como requisito parcial para optar al título de:

Magister en Ingeniería - Analítica

Director (a):

Ph.D. Francisco Javier Moreno Arboleda

Línea de Investigación:

Minería de datos

Universidad Nacional de Colombia

Facultad de Minas, Departamento de Ciencias de la Computación y de la Decisión

Medellín, Colombia

2022

Dedicatoria

A mis padres Blanca Elena Correa y Guillermo Arboleda.

Declaración de obra original

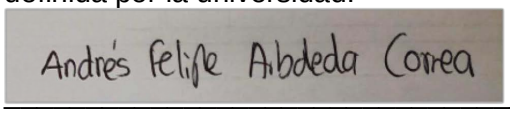
Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.



Nombre

Fecha 31/07/2022

Agradecimientos

Al profesor Francisco Moreno por su mentoría y apoyo en la conceptualización, construcción y desarrollo de esta tesis.

Resumen

Propuesta para múltiples diseños de tiendas basados en reglas de asociación

El éxito de una tienda depende de su capacidad para comprender el comportamiento de los clientes y su capacidad para reaccionar ante cambios en los hábitos de consumo. El análisis de datos sobre las compras que hacen los clientes es clave para buscar nuevas oportunidades de ventas. En este trabajo, se presenta el estado del arte basado en las directrices para la revisión y clasificación sistemática de la literatura, y se propone un algoritmo para la generación de múltiples diseños de las secciones de una tienda en la cual se considera: i) la minería de reglas de asociación, ii) el número total de unidades compradas de los productos en las transacciones en la generación de las reglas (aspecto que suele ser ignorado en las reglas de asociación clásicas), iii) una estructura jerárquica para la clasificación de productos desarrollada por el Departamento de Asuntos Económicos y Sociales de las Naciones Unidas, iv) la selección de reglas de asociación interesantes (que cumplen ciertos umbrales) y v) un conjunto de restricciones que establecen que ciertos productos o categorías de productos no deben estar cercanos en una sección de la tienda. Finalmente, se presenta un experimento con la generación de diseños para el cual se aplica el algoritmo propuesto sobre un conjunto de datos públicos de ventas de un supermercado con el fin de ver la viabilidad de la propuesta.

Palabras clave: minería de datos; análisis de cestas de compra; reglas de asociación; diseño de tiendas.

Abstract

Proposal for multiple store layouts based on association rules

The success of a store depends on its ability to understand the behavior of its customers and its ability to react to changes in consumer habits. The analysis of customer purchases is key to finding new sales opportunities. In this paper, we present the state of the art based on the guidelines for systematic literature review and classification, and propose an algorithm for the generation of multiple layouts of the sections for a store in which we consider: (i) association rule mining, (ii) the total number of units purchased of the products in the transactions in the generation of the rules (an aspect that is usually ignored by classical association rules), (iii) a hierarchical structure for product classification developed by the United Nations Department of Economic and Social Affairs, (iv) the selection of interesting association rules (that meet certain thresholds), and (v) a set of constraints that state that certain products or categories of products should not be placed close in a section of a store. Finally, an experiment with the generation of designs is presented by applying the algorithm on a set of public supermarket sales to check the feasibility of the proposal.

Keywords: data mining; market basket analysis; association rules; store design.

Contenido

Pág.

Resumen ..... **IX**

Lista de figuras ..... **XIII**

Lista de tablas ..... **XIV**

Lista de Abreviaturas ..... **XVI**

Introducción **1**

1. Objetivos **5**
1.1 Objetivo general 5
1.2 Objetivos específicos 5

2. Metodología del estado del arte **7**
2.1 Definición de la pregunta de investigación 7
2.2 Desarrollo de la investigación 7
2.3 Selección de documentos relevantes 8
2.4 Clasificación basada en el resumen 8
2.5 Proceso de extracción de datos y clasificación 9

3. Estado del arte **11**
3.1 Reglas de asociación donde se considera la cantidad 11
3.2 Diseño de instalaciones mediante reglas de asociación 13
3.3 Reglas de asociación en los supermercados 14
3.4 Enfoques de diseño de supermercados 15

4. Líneas de investigación **21**
4.1 Diseño de bodegas para optimizar la ubicación de los productos de supermercados 21
4.2 Diseño de bodegas optimizadas para recorridos de operarios robots 22

5. Planteamiento del problema **25**
5.1 Ejemplo ilustrativo 25

6. Algoritmo **35**
6.1 Entradas 35
6.2 Paso 1: Reglas de asociación y la cantidad 37
6.2.1 Paso 1.1: Minería de reglas de asociación mediante el algoritmo Apriori. 38
6.2.2 Paso 1.2: Cálculo de la medida Wgain 39

6.2.3 Paso 1.3: Cálculo de la medida *Ugain* 40
6.2.4 Paso 1.4: Cálculo de *UWscore* a partir de *Wgain* y *Ugain*. 42
6.2.5 Paso 1.5: Selección de reglas de asociación significativas a partir de *UWscore*. 43
6.3 Paso 2: Selección de reglas con *support*, confidence y *lift* significativos 46
6.4 Paso 3: Selección de reglas que cumplen las restricciones 48
6.5 Paso 4: Generar grupos de productos cercanos 50
6.6 Paso 5: Generar diseños 58
6.7 Resumen 59

7. Experimento 63
7.1 Paso 1: Reglas de asociación y la cantidad 64
7.1.1 Paso 1.1: Minería de reglas de asociación mediante el algoritmo Apriori 64
7.1.2 Paso 1.2: Cálculo de la medida Wgain 66
7.1.3 Paso 1.3: Cálculo de la medida *Ugain* 67
7.1.4 Paso 1.4: Cálculo de *UWscore* a partir de *Wgain* y *Ugain*. 67
7.2 Paso 2: Selección de reglas con *support*, confidence y *lift* significativos 68
7.3 Paso 3: Selección de reglas que cumplen las restricciones 69
7.4 Paso 4: Generar grupos de productos cercanos 69
7.5 Paso 5: Generar diseños 70

8. Conclusiones y resultados 73
8.1 Resultados 74

A. Anexo: Transacciones experimento. 77

B. Anexo: Categorización de los productos. 80

C. Anexo: Resultados de *WGain*, *UWScore* reglas de asociación 86

Bibliografía 89

Lista de figuras

Pág.

Figura 2-1: El proceso de clasificación sistemática. Adaptado de [9]. 7
Figura 2-2: El proceso de clasificación de los artículos. Adaptado de [9]. 8
Figura 3-1: Diseño de un supermercado derivado de escalamiento multidimensional. Fuente [1]. 14
Figura 3-2: Escalamiento dimensional de las relaciones entre productos. Fuente [2]..... 16
Figura 5-1: Árbol de clasificación de productos. Adaptado de [35]. 26
Figura 6-1: Reglas de asociación en su correspondiente *cluster*. 51
Figura 6-2: Dendrograma primera iteración. 54
Figura 6-3: Dendrograma segunda iteración. 57
Figura 6-4: Dendrograma con cuatro *clusters*. 57
Figura 6-5: Dendrograma completo. 58
Figura 6-6: Dendrograma con cuatro *clusters* calculado a partir del dendrograma completo de la Figura 6-5. 58
Figura 6-7: Propuesta de diseño de tienda. 59
Figura 6-8: Resumen de los pasos del algoritmo. 60
Figura 7-1: Fragmento de código en Python para la generación de números enteros pseudoaleatorios. 66
Figura 7-2: Dendrograma de los *clusters* obtenidos. 69
Figura 7-3: Propuesta de diseño para el supermercado con $g = 7$. 70
Figura 7-4: Propuesta de diseño para el supermercado con $g = 4$. 71

Lista de tablas

Pág.

Tabla 3-1: Comparación de los trabajos analizados. 18
Tabla 5-1: Muestra de transacciones (compras). 25
Tabla 5-2: Algunas estadísticas de los productos de la Tabla 5-1 clasificados según el árbol de clasificación de la Figura 5-1**.** 27
Tabla 5-3: Primera iteración Apriori: generación de *frequent itemsets.* 31
Tabla 5-4: Segunda iteración Apriori: generación de *frequent itemsets.* 32
Tabla 5-5: Tercera iteración Apriori: generación de *frequent itemsets.* 33
Tabla 5-6: Reglas de asociación generadas. 34
Tabla 6-1: Reglas de asociación ordenadas descendentemente por *confidence* generadas de las transacciones de la Tabla 5-1. 38
Tabla 6-2: *Wgain* de las primeras 5 reglas de asociación de la Tabla 6-1. 40
Tabla 6-3: *Ugain* de las primeras 5 reglas de asociación de la Tabla 6-1. 41
Tabla 6-4: *UWScore* de las cinco mejores reglas de asociación de la Tabla 6-1. 43
Tabla 6-5: Reglas de asociación de Tabla 6-1 que cumplen con el umbral de $UWScore \geq 14$. 44
Tabla 6-6: Reglas de asociación resultantes de la Tabla 6-5 que cumplen con los umbrales $minsupp = 0.1$, $minconf = 0.5$ y $minlift = 0.8$. 46
Tabla 6-7: Reglas de asociación resultantes de la Tabla 6-6 que cumplen con las restricciones $RE = RE1, RE2, RE3, RE4, RE5, RE6, RE7$ 49
Tabla 6-8: Matriz P de distancias calculadas entre los pares de *cluster*. 53
Tabla 6-9: Matriz de distancias con datos faltantes para el nuevo *cluster* $R1$,$R30$. 55
Tabla 6-10: Algunas fórmulas de *linkage* para calcular la distancia entre *cluster*. Adaptada de [49] [50]. 55
Tabla 6-11: Matriz de distancias para el nuevo *cluster* $R1$,$R30$. 56
Tabla 7-1: Muestra de 10 transacciones. 63

Tabla 7-2: Muestra de la clasificación de los productos según la propuesta de [35]..... 64
Tabla 7-3: Reglas de asociación ordenadas descendentemente por *confidence* generadas a partir de las transacciones del Anexo A..... 64
Tabla 7-4: Cantidad total obtenida para cada *group*..... 66
Tabla 7-5: *Wgain* de las primeras 5 reglas de asociación de la Tabla 7-3..... 66
Tabla 7-6: *UWScore* de las cinco mejores reglas de asociación de la Tabla 7-3. 67
Tabla 7-7: Reglas de asociación resultantes de la Tabla 7-7 que cumplen con los umbrales $minsupp = 0.1$, $minconf = 0.4$ y $minlift = 1$..... 68

Lista de Abreviaturas

Abreviaturas

Abreviatura	Término
Ref	Referencia
MDS	*Multidimensional scaling*
MBA	*Market basket analysis*
Q-TID	*Quantity-transaction identifiers*
FP-Growth	*Frequent pattern growth*
QBARM	*Quantity based association rule mining*
LHS	*Left hand side*
RHS	*Right hand side*
MINWAL	*Mining weighted algorithm*
WARM	*Weighted association rule mining*
BWARM	*Binary WARM*

Introducción

El diseño de un supermercado es un elemento estratégico para aumentar o disminuir sus ventas, ya que puede llegar a influir en el comportamiento de compra de los clientes y en la eficiencia operativa. Un supermercado se suele rediseñar frecuentemente para afrontar las necesidades cambiantes del mercado, e.g., ingreso de nuevos productos y marcas, cambios en las preferencias de compras de los clientes, políticas de mercadeo (promociones, eventos en fechas especiales), entre otras. Un diseño puede incitar o incluso obligar a un cliente a pasar por una determinada área del supermercado lo que puede incrementar la posibilidad de que este compre más [1]. En general, el objetivo de un supermercado es la maximización del beneficio. El objetivo es que los clientes compren el mayor número posible de productos.

Borges et al. [2] expresa que el diseño de un supermercado es una tarea demandante. Parte de la complejidad radica en la relación entre las categorías en venta, en el impacto en el comportamiento del consumidor y en el tráfico en el supermercado. En el diseño de un supermercado no solo se propende por aprovechar el espacio, sino que también interesa aumentar las ventas: los estantes deben atraer la atención del cliente [1]; además, se debe considerar su perfil: estantes dirigidos a los niños, a las mujeres jóvenes, a los adultos mayores, entre otros. La disposición de los departamentos (secciones), que tienen productos o categorías de productos relacionados, cerca unos de otros tienen un impacto considerable en la rentabilidad de los supermercados [1].

Existe una gran cantidad de posibilidades de disposición de los productos dentro de un supermercado. Por un lado, las disposiciones de las estanterías pueden ser muy variadas; además, las disposiciones de los productos en los estantes (es decir, en qué posición se coloca un producto en un estante determinado) también pueden ser muy diversas.

En el abordaje del problema del gerente de rediseñar las secciones de los productos se usa la minería de datos y las reglas de asociación para encontrar cuáles productos o cuáles categorías de productos se deberían ubicar cerca (en la misma sección). La minería de reglas de asociación examina una base de datos de transacciones (compras) para determinar cuáles artículos se compran con más frecuencia de forma conjunta [1] [3]. El objetivo es extraer correlaciones, patrones frecuentes, asociaciones o estructuras causales entre conjuntos de productos en las bases de datos [1] [4].

El sector minorista es una de las principales áreas de aplicación de la minería de datos, ya que recoge enormes cantidades de datos sobre las ventas y el historial de compras de los clientes. Varios estudios se han centrado en el desarrollo de estrategias de mercadeo como el mercadeo orientado al cliente [1] [5].

Una regla de asociación [6] es una forma de medir la relación entre los productos a partir de una base de datos de transacciones. Los análisis basados en la minería de reglas de asociación se han llevado a cabo en una gran variedad de conjuntos de datos y son especialmente útiles en el análisis de grandes conjuntos de datos [7].

El algoritmo Apriori descubre los conjuntos de elementos (ítems, productos) frecuentes considerando la ocurrencia del conjunto de elementos en las transacciones y no suele considerar aspectos como el beneficio o la cantidad. Sin embargo, en un sistema de compras puede ocurrir que algunos productos cumplan tres condiciones: i) aparezcan solo en unas pocas transacciones, ii) sean comprados en grandes cantidades y iii) sean determinantes para las ganancias del negocio. Por lo tanto, estos productos son decisivos para las ventas.

En esta tesis se propone un algoritmo para la generación de múltiples diseños para una tienda basado en las reglas de asociación. Para ello se considera: i) la minería de reglas de asociación, ii) el número total de unidades compradas de los productos en las transacciones para la generación de las reglas (aspecto que suele ser ignorado en las reglas de asociación clásicas), iii) una estructura jerárquica para la clasificación de productos desarrollada por el Departamento de Asuntos Económicos y Sociales de las Naciones Unidas, iv) la selección de reglas de asociación interesantes (que cumplen ciertos umbrales) y v) un conjunto de restricciones que establecen que ciertos productos

o categorías de productos no deben estar cercanos (en una sección de la tienda). En el desarrollo del algoritmo se presenta un ejemplo ilustrativo.

En la metodología de investigación se siguen las directivas para la revisión sistemática [8] y del estudio del proceso para la clasificación sistemática [9], para las cuales se usa un protocolo de búsqueda en dos bases de datos científicas para identificar, extraer y analizar las publicaciones relevantes. A partir de lo anterior se desarrolla y se presenta el estado del arte.

Esta tesis está organizada así: en la sección 1) se presenta el objetivo general y los objetivos específicos, en la sección 2) la metodología del estado del arte siguiendo los lineamientos de revisión y clasificación sistemática, en la sección 3) se presenta el estado del arte, en la sección 4) se presentan las líneas de investigación sobre el problema de la ubicación de productos, en la sección 5) se plantea el problema mediante un ejemplo ilustrativo y se presentan algunas definiciones de las reglas de asociación, en la sección 6) se propone un algoritmo para solucionar el problema y se aplica sobre el ejemplo ilustrativo, en la sección 7) se presenta un experimento con la generación de diseños donde se aplica el algoritmo sobre un conjunto de datos públicos de ventas de un supermercado con el fin de analizar los resultados de la propuesta y finalmente en 8) se concluye y se presentan los resultados obtenidos.

1. Objetivos

1.1 Objetivo general

Proponer un método para la generación de múltiples diseños para una tienda basado en reglas de asociación.

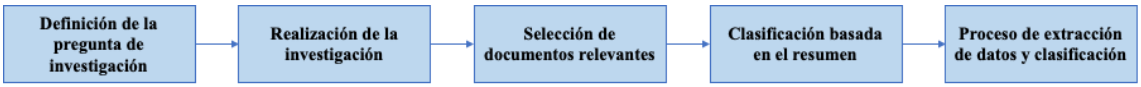
1.2 Objetivos específicos

- Identificar algoritmos que consideren el diseño de una tienda y los criterios que aplican para dicho diseño.
- Proponer un método o extender alguno de los identificados en el objetivo anterior para el diseño de una tienda generando múltiples diseños.
- Validar el método propuesto o extendido mediante un prototipo funcional.

2. Metodología del estado del arte

En la elaboración del estado del arte, se adoptan las directrices para la revisión sistemática de la literatura propuestas por Kitchenham & Charters [8] y del estudio del proceso para la clasificación sistemática propuesto por Petersen et al. [9]. Como se explica en Petersen et al. [9] el objetivo de un estudio de clasificación sistemática es obtener una visión general del área de investigación y complementarlo con la investigación del estado de evidencias. Para esta tesis los resultados ayudan a identificar cómo se han usado las reglas de asociación en el diseño de supermercados y determinar que aplicaciones se han desarrollado. También ayuda a identificar posibles oportunidades de investigación, identificar que tipos de enfoque se han usado y problemas por resolver. En la Figura 2-1 se presentan los pasos del proceso.

Figura 2-1: El proceso de clasificación sistemática. Adaptado de [9].



2.1 Definición de la pregunta de investigación

En el primer paso, se define como pregunta de investigación: ¿es posible generar diseños para un supermercado a partir de la información de las transacciones (compras) basado en reglas de asociación?

2.2 Desarrollo de la investigación

En el segundo paso, los principales documentos para el estudio se seleccionan de las bases de datos científicas mediante cadenas de búsqueda o palabras claves En la búsqueda se incluyeron bases de datos científicas como IEEE Xplore y ScienceDirect. Se analizaron artículos publicados en revistas indexadas y conferencias. Se usaron palabras claves como: *association rules*, *association rules in supermarket design*, *association*

rules and item quantity, *design of a facility*, entre otras. La búsqueda bibliográfica se llevó a cabo sin restricción de tiempo, debido a que el tema es relativamente nuevo y; por lo tanto, toda la literatura de esta área se considera relevante para el estudio.

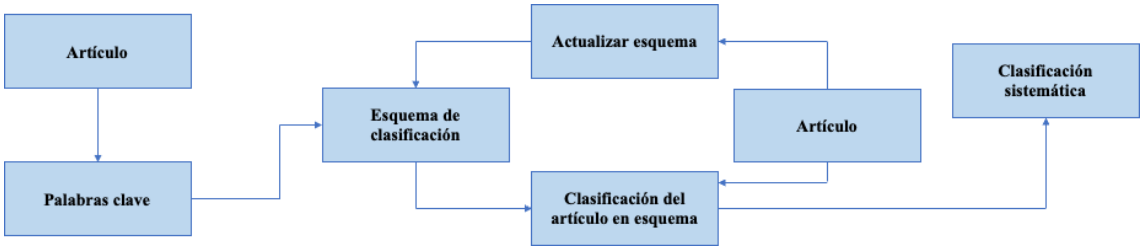
2.3 Selección de documentos relevantes

Una vez obtenidos los artículos según la búsqueda hecha, el siguiente paso fue examinar su relevancia. La primera etapa fue examinar la pertinencia de los artículos en función de sus títulos, se descartaron los artículos recuperados cuyos títulos indicaban que no eran relevantes para el análisis. La segunda etapa fue la lectura de los resúmenes de los artículos que habían pasado la primera etapa, en algunos casos fue necesario leer la introducción y sus conclusiones para determinar si el artículo era relevante para el estudio.

2.4 Clasificación basada en el resumen

En este paso del proceso se pretende clasificar los artículos relevantes en categorías, para esto se siguió el proceso definido por Petersen et al. [9] como se muestra en la Figura 2-2. Este proceso extrae de los resúmenes de los artículos algunas palabras clave y conceptos que reflejen las contribuciones. A partir de esto se agrupan en diferentes categorías, e.g., diseño de supermercados, diseño de tiendas, reglas de asociación, entre otras. Cada artículo se leyó en su totalidad y si el contenido pertenecía a una categoría diferente se hacía la actualización correspondiente. El resultado de este proceso es la clasificación de todos los artículos en las categorías.

Figura 2-2: El proceso de clasificación de los artículos. Adaptado de [9].



2.5 Proceso de extracción de datos y clasificación

En este último paso se extrajo la información requerida de los artículos de investigación con el fin de analizar e identificar posibles oportunidades de investigación con el fin de presentar problemas por resolver en área. En la sección 3 se presenta el estado del arte construido a partir de los artículos seleccionados y en la sección 4 se presentan problemas por resolver.

3. Estado del arte

En esta sección se presentan los trabajos enfocados en las reglas de asociación donde se considera la cantidad de unidades compradas de cada producto, el diseño de instalaciones usando reglas de asociación, las reglas de asociación en los supermercados y enfoques de diseño de supermercados. Finalmente, se resumen y se comparan los trabajos.

3.1 Reglas de asociación donde se considera la cantidad

El algoritmo de minería de reglas de asociación clásico descubre los conjuntos de elementos frecuentes considerando la ocurrencia del conjunto de elementos en las transacciones y no otros aspectos como el beneficio o la cantidad.

En un sistema de compras puede ocurrir que algunos productos cumplan tres condiciones: i) aparezcan solo en unas pocas transacciones, ii) sean comprados en grandes cantidades y iii) sean determinantes para las ganancias del negocio. Por lo tanto, estos productos son decisivos para las ventas.

En [10] se indica que en las cestas de compra, un mismo producto puede aparecer varias veces en una misma transacción. Así, dado un producto, o de forma más general un conjunto de productos, se puede considerar el número total de transacciones que contienen ese conjunto (sin considerar repeticiones del conjunto en una misma transacción) y el número total de veces que aparece ese conjunto en todas las transacciones (i.e., acá sí se cuentan las posibles repeticiones del conjunto). Se propone un marco de trabajo para obtener las reglas de asociación, mediante el algoritmo Apriori, considerando estos dos casos. Se concluye que las reglas son efectivas para estudiar asociaciones entre productos para este tipo de cestas de compra, a diferencia del

enfoque tradicional en el que se considera que un conjunto de productos solo ocurre una vez en cada transacción.

En [11] se propone un sistema para analizar y pronosticar las ventas en un supermercado. Se obtienen reglas de asociación mediante el algoritmo Apriori, considerando los productos comprados en cada transacción y su cantidad. Esta información puede ayudar a descubrir la evolución (crecimiento, estabilidad, decrecimiento) de las ventas de cada producto. Sin embargo, no se explica cómo se aplicó la cantidad comprada de cada producto para determinar las reglas. Tampoco se presenta el grado de acierto de las predicciones.

En [12] se propone un algoritmo basado en un árbol Q-TID (*quantity-transaction identifiers*) para descubrir conjuntos de productos frecuentes basados en la cantidad y se compara con los algoritmos FP-Growth (*frequent pattern growth*) [13], [14] y Apriori. Además de considerar la cantidad y el costo, el algoritmo propuesto presentó un menor tiempo de ejecución que los otros dos: requirió cerca de un 70% menos del tiempo que el Apriori y menos del 60% que el FP-Growth.

En [15] se propone un algoritmo de reglas de asociación basado en la cantidad (QBARM*, Quantity Based Association Rule Mining*). Se considera y pondera la frecuencia de las apariciones de los elementos en las transacciones. Este algoritmo es una extensión al presentado en [16] ya que considera la cantidad ponderada, siendo esta su principal novedad.

En [17] se propone un algoritmo basado en la cantidad ponderada y en la utilidad de las reglas de asociación significativas. Inicialmente, se usa el algoritmo Apriori para generar un conjunto de reglas de asociación a partir de una base de datos y a continuación, para cada regla de asociación generada, se calcula una puntuación ponderada de utilidad combinada (*UW-Score*) basada en las restricciones de ponderación (*W-gain*, cantidad en la que un elemento es adquirido) y en la utilidad (*U-gain*). Por último, se determina un subconjunto de reglas de asociación en función de la puntuación *UW*.

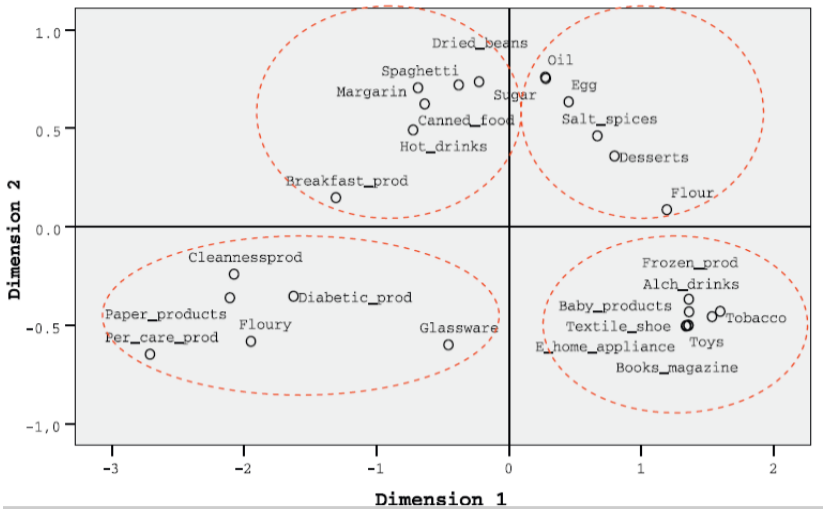
3.2 Diseño de instalaciones mediante reglas de asociación

En [18] se aplica *Market Basket Analisys* (MBA) para diseñar las instalaciones de una sala de juegos *arcade*. Se consideran dos propuestas para el diseño. El primer diseño aplicó la teoría sensacional de entrada (*entrance sensational theory*) [19], este diseño también considera la frecuencia de juego de cada categoría donde los juegos de categoría de premios se colocan en las esquinas debido a que tienen la mayor frecuencia de juego. El propósito de esta consideración es que los clientes estarán alrededor de otros juegos cuando estén buscando un juego de premios. En este diseño se agrupan los juegos según sus categorías (juegos de apuestas (*gambling*), de simulación, de cartas, para niños) para aumentar la probabilidad de que se continúen jugando juegos de la misma categoría. Las reglas obtenidas son a nivel de juegos, e.g., $\{ORECA\ BATTLE\} \Rightarrow \{HERO\ ROBOTS\}$. Una ventaja de este diseño es que los clientes que les guste los juegos de una categoría pueden conocer otros juegos de la misma categoría.

El segundo diseño es similar al primero, pero se aplica MBA entre categorías. Las reglas obtenidas son a nivel de categorías de juegos, e.g., $\{Simulation\} \Rightarrow \{Gambling\}$. En este diseño, el juego de ingresos más bajos de una categoría se coloca cerca del juego de ingresos más altos de otra categoría. De esta forma, si una persona juega un juego de una categoría es muy probable que aumente los ingresos de los juegos con ingresos más bajos. Para los dos diseños propuestos no se detallan los algoritmos para obtener las reglas ni se analiza si se obtuvieron mejores ingresos con las propuestas.

En [1] se propone el diseño de supermercados con base en la asociación entre categorías de productos mediante reglas de asociación (se usa el algoritmo Apriori) y escalamiento multidimensional (MDS). El escalamiento multidimensional permite representar visualmente la proximidad (ya sea por similitud o por distancia) entre un conjunto de objetos (aquí, categorías de productos). De esta forma, se genera el diseño para la distribución de los productos en un supermercado con base en las categorías. En la Figura 3-1 se muestra un ejemplo de un diseño.

Figura 3-1: Diseño de un supermercado derivado de escalamiento multidimensional. Fuente [1].



Se concluye que el diseño es ventajoso ya que: a) considera lo que los clientes quieren (al generar las reglas de asociación a partir de las transacciones del supermercado) y b) satisface sus necesidades de una forma más eficiente (al considerar la proximidad de las categorías).

3.3 Reglas de asociación en los supermercados

En [20] se propone un método para descubrir patrones de compra considerando múltiples supermercados. Dos de las debilidades de las reglas de asociación son que suponen que i) los patrones no cambian a lo largo del tiempo y ii) que todos los productos siempre están disponibles en los estantes del supermercado (esto no ocurre, e.g., con los productos de temporada). Además, se debe considerar que algunos productos no se venden en algunos supermercados, e.g., por razones geográficas, económicas o culturales.

Para resolver este problema se propone un método llamado *Store chain association rules*, donde las diferentes tiendas pueden tener diferentes estrategias con respecto a los productos que se compran juntos. Se propone un algoritmo similar al Apriori para obtener las reglas. Las reglas contienen además información de la tienda (su ubicación) y el

tiempo cuando estas se cumplen. Por ejemplo, una regla puede indicar: "*En enero, los clientes compran medicamentos para el resfriado, humidificadores, café y gafas de sol juntos en los supermercados cercanos a las estaciones de esquí*". Concluyen que estas reglas generadas tienen una ventaja sobre las generadas con Apriori debido a que tienen información adicional que pueden ser aplicadas en estrategias de mercadeo y en distribución para toda la cadena de tiendas.

En [21] se compara el número total de reglas y la cardinalidad de sus conjuntos de productos mediante los algoritmos Apriori y K-Apriori [22]. El algoritmo K-Apriori generó reglas de hasta cinco productos con un 20% de soporte. El algoritmo Apriori solo generó reglas de dos productos con 5% de soporte y de tres productos con 4% de soporte. Se afirma que con los resultados del algoritmo K-Apriori para la tienda estudiada se mejoraron sus ventas, aunque no se especifica en cuánto mejoraron.

En [23] se analizaron los datos de ventas de un supermercado, se buscaron reglas y patrones entre diferentes grupos de productos para aplicarlas en un enfoque orientado al cliente. Se obtuvieron 156 reglas con el algoritmo Apriori. Se concluyó que estas reglas se pueden aplicar en el diseño de una tienda, de tal manera que los productos que suelen ser adquiridos juntos por los clientes estén ubicados de tal forma que el cliente recuerde su ubicación. Sin embargo, los autores no desarrollan la idea del diseño de un supermercado con las reglas obtenidas.

3.4 Enfoques de diseño de supermercados

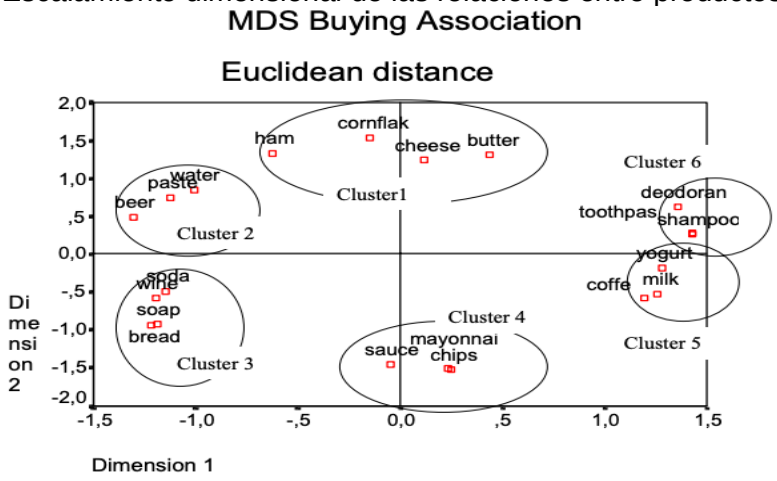
Las tiendas minoristas se organizan tradicionalmente en departamentos (áreas), i.e., en un departamento se juntan los productos que comparten algunas características funcionales, e.g., productos con el mismo propósito (e.g., aseo, entretenimiento) o del mismo origen (productos del mar). Este enfoque [24] no es adecuado para responder a las necesidades de los compradores que tienen prisa [1]. Sin embargo, no se aclara porque no es adecuado, aunque quizás es porque un comprador que tiene prisa debe recorrer gran parte de la tienda para reunir los productos básicos que requiere.

En el enfoque propuesto en [2] se agrupan los productos en torno a *universos de consumo* relacionados con la asociación de compras. Así, en lugar de encontrar café en

la sección de bebidas, queso en la sección de lácteos, jamón en la sección de carnes y pan en la sección de panadería, se podrían encontrar esos cuatro productos en el universo de consumo de desayuno. Otros universos, como el universo del bebé o el universo de la vajilla, proponen el mismo esquema para agrupar diferentes categorías de productos. Este enfoque se basa en el principio de que el uso conjunto de productos producirá compras conjuntas.

Se obtuvieron reglas de asociación a partir de 1.700.000 transacciones con productos clasificados en diferentes categorías, e.g., agua, queso, mantequilla. Se encontró la presencia correlacionada, i.e., la probabilidad de encontrar un producto B dado que el producto A está en una canasta de compra. Se calculó una matriz de correlación y finalmente se aplicó una técnica de MDS para representar espacialmente estas relaciones. En la Figura 3-2 se representan las categorías en el espacio multidimensional.

Figura 3-2: Escalamiento dimensional de las relaciones entre productos. Fuente [2].



En [25] se analizan los datos de las rutas que siguen los compradores de un supermercado mediante algoritmos de agrupamiento multivariante, que corresponden a análisis de datos que contienen observaciones con más de una variable medida [26]. Los cuales pueden ser aplicados a diseños con una gran variedad de restricciones espaciales. El objetivo es comprender el comportamiento de los compradores dentro de la tienda sin considerar las tácticas de comercialización. Se identificaron 14 grupos típicos de rutas. Concluyen que estas rutas pueden ser de ayuda para el diseño de un

supermercado.

En [27] se propone un enfoque de minería de datos para el diseño de un supermercado. Se propone un algoritmo HUIM (*high-utility itemset mining*) basado en las utilidades de las ventas para reorganizar la tienda y determinar las relaciones entre las familias de productos (aunque no dan ejemplos de estas). Las cantidades y los precios de los artículos correspondientes a las familias de productos son considerados en este enfoque. El enfoque tiene dos etapas: En la primera se aplica el algoritmo HUIM para determinar las asociaciones entre artículos según las compras. En la segunda etapa se aplica un algoritmo de reglas de asociación ponderadas [28] para determinar el diseño del supermercado considerando las asociaciones encontradas para ubicar los productos que suelen ser comprados juntos. Se concluye que el diseño de una tienda no es estático ya que se construye a partir de los hábitos de consumo de los clientes y puede ser usado por los decisores para asegurar la sostenibilidad de las tiendas.

En [28] se aplican diferentes técnicas de minería de datos basados en reglas de asociación ponderadas para problemas de diseño de instalaciones. Las técnicas son MINWAL(O) (*mining weighted algorithm*), MINWAL(W), WARM (*weighted association rule mining*) y BWARM (*binary* WARM). En las dos primeras no se aclara el significado de la O y la W. A cada elemento (producto) se le asigna una ponderación de acuerdo con factores como: i) la demanda, ii) el factor de manipulación de piezas y iii) la eficiencia del equipo de manipulación de los materiales, los cuales son considerados los más importantes en el problema de diseño de instalaciones. Concluyen que este estudio se diferencia de los demás debido a que considera estos tres factores juntos y obtienen como mejor enfoque el uso de WARM ya que obtuvieron mejores resultados en términos de la cantidad de productos producidos, tiempos de ciclos y tiempos de espera en cola.

En la Tabla 3-1 se comparan los trabajos analizados en esta sección.

Tabla 3-1: Comparación de los trabajos analizados.

Ref	Enfoque de las reglas de asociación	Propuesta	Observaciones
[10]	No	Un marco de trabajo para obtener reglas de asociación mediante el algoritmo Apriori considerando la cantidad.	Considera la cantidad para calcular las reglas de asociación.
[11]	Tendencias de ventas y pronóstico de cantidad	Un sistema para analizar las tendencias de las ventas de un supermercado y el pronóstico de las cantidades de sus productos mediante el algoritmo Apriori.	Puede ayudar a descubrir la evolución (crecimiento, estabilidad, decrecimiento) de las ventas de los productos. Considera la cantidad para calcular las reglas de asociación.
[12]	No	Un algoritmo basado en un árbol Q-TID para descubrir conjuntos de productos comprados frecuentemente.	El algoritmo propuesto requirió cerca de un 70% menos del tiempo de ejecución que el Apriori y menos del 60% que el FP-Growth. Considera la cantidad para calcular las reglas de asociación.
[15]	No	Un algoritmo QBARM de reglas de asociación basado en la cantidad.	Considera la cantidad para calcular las reglas de asociación.
[17]	No	Un algoritmo que genera reglas de asociación significativas basadas en la cantidad y la utilidad.	Se obtienen reglas de asociación mediante el algoritmo Apriori. Cada regla se pondera y se seleccionan las reglas significativas.
[18]	Diseño de salas de juego	Aplicación de MBA para diseñar salas de juegos arcade.	Se proponen dos diseños para una sala de juego considerando la frecuencia de juego de las categorías de los juegos.
[1]	Diseño de un supermercado	Diseño de supermercados con base en la asociación entre categorías de productos mediante reglas de asociación y MDS [2].	Para el diseño de un supermercado se considera la proximidad de las categorías. Solo se propone un diseño.
[20]	Ambiente de múltiples tiendas	Método para descubrir patrones de compra considerando múltiples tiendas.	Las reglas de asociación obtenidas incluyen su tiempo de validez y la ubicación de la tienda. No se aplican las reglas obtenidas para el diseño de una tienda.
[21]	Análisis de los productos de un supermercado	Se compara el número total de reglas y la cardinalidad de los conjuntos de productos obtenidos en ellas mediante los algoritmos Apriori y K-Apriori	Presenta un caso de estudio que permite evaluar los algoritmos Apriori y K-Apriori.
[23]	Aproximación orientada	Se buscan reglas y patrones entre diferentes productos de	Se obtienen prácticas proactivas orientadas al cliente, con las que

	al cliente para un supermercado	un supermercado para aplicarlas en un enfoque orientado al cliente.	se pretenden optimizar las ventas y la satisfacción del cliente, logrando una ventaja competitiva y sostenible. No se desarrolla la idea del diseño de un supermercado con las reglas obtenidas.
[24]	Diseño de tienda. No se usan reglas de asociación	Distribución de una tienda en departamentos en los que se reúnen los productos que comparten algunas características funcionales o del mismo origen.	Es un diseño estático ya que no considera el comportamiento de las ventas.
[2]	Diseño de un supermercado	Se generan reglas de asociación y se aplica el *MDS* para proponer el diseño de un supermercado.	Se agrupan diferentes categorías de productos basado en el principio de que el uso conjunto de productos producirá compras conjuntas. Solo se propone un diseño para un supermercado.
[25]	Búsqueda exploratoria de rutas de compras. No se usan reglas de asociación	Se pretende identificar grupos típicos de rutas para comprender el comportamiento de los compradores en una tienda.	Se identificaron 14 grupos típicos de rutas. No se propone un diseño de la tienda basado en tales rutas.
[27]	Diseño de un supermercado	Un enfoque de minería de datos para el diseño de una tienda. Se propone un algoritmo *HUIM,* basado en las utilidades de las ventas para reorganizar la tienda y determinar las relaciones entre las familias de productos.	Se consideran las utilidades de los productos. Solo se propone un diseño para un supermercado.
[28]	Diseño de una instalación	Diseño de instalaciones y verificación de su viabilidad mediante dos casos de estudio.	Se consideran la demanda, el factor de manipulación de piezas y la eficiencia del equipo de manipulación de los materiales para el diseño de las instalaciones. Se pueden aplicar en problemas de diseño de instalaciones con múltiples objetivos [29]. Solo se propone un diseño para una instalación.

4. Líneas de investigación

En esta sección se exploran problemas por resolver en la ubicación de productos en diferentes espacios como supermercados y bodegas. Estos problemas están relacionados con su diseño y distribución.

4.1 Diseño de bodegas para optimizar la ubicación de los productos de supermercados

En el diseño de una bodega interviene un gran número de factores interrelacionados que afectan el coste de su instalación. Además, transcurrido un tiempo, el diseño se puede volver ineficiente debido a factores como las tendencias del mercado y los cambios en las estrategias de la empresa, por lo que puede ser necesario un nuevo diseño. En un diseño se deben analizar los requisitos del sistema, como los costes de inversión y los operativos, la capacidad de almacenamiento, el tiempo de respuesta y la calidad en el cumplimiento de los pedidos, entre otros. A continuación, se presentan algunos trabajos que aluden a este problema.

En [30] se analiza el diseño de una bodega y la disposición de los artículos para la preparación de pedidos. Se analizó un frigorífico que almacena un amplio abanico de productos como leche, queso, mermeladas, miel, yogur, mantequilla, postres, chocolate, aceite líquido, helados y alimentos para bebés. Se analizó su diseño, el método para la disposición de los artículos y las políticas de preparación de los pedidos. Los resultados de los distintos escenarios se evaluaron mediante simulación. Se propone un nuevo diseño para el frigorífico y decisiones de disposición de los artículos para hacer más productiva la preparación de los pedidos. Para esto se propone un modelo matemático y se aplica MDS. Se concluye que con el diseño propuesto se reduce el tiempo de preparación de pedidos promedio de 1.542 minutos a 0.66 minutos.

En [31] se propone un método heurístico basado en algoritmos genéticos para resolver el problema de asignación de pedidos (SAP, *storage assignment problem*), un problema de dificultad NP, i.e., encontrar una forma eficaz de ubicar los productos en un almacén para mejorar la eficiencia operativa en la preparación de los pedidos. Los resultados que obtienen indicaron que el rendimiento con el método propuesto es estadísticamente mejor que al usar una política aleatoria.

4.2 Diseño de bodegas optimizadas para recorridos de operarios robots

El diseño de un almacén o bodega desempeña un papel fundamental en la mejora de la eficiencia del transporte. Esta labor es hecha casi siempre por humanos, lo que en ocasiones conduce a resultados subóptimos. Debido al gran crecimiento de la industria de los envíos, las bodegas inteligentes que emplean robots autónomos para transportar paquetes se han comenzado a usar ampliamente en la última década para optimizar este proceso. A continuación, se presentan algunos trabajos que tratan este problema.

En [32] se presenta un método basado en algoritmos evolutivos para diseñar automáticamente la distribución de bodegas para un entorno en el cual robots transportan paquetes. Para hacer frente a la eficiencia de cada diseño, se entrena una red neuronal para predecir los resultados de las distribuciones.

En [33] se propone una solución basada en inteligencia evolutiva (EVIN, *evolutionary intelligence*) para el diseño de sistemas de gestión de bodegas. Una unidad analítica basada en una red neuronal permite predecir el número óptimo de robots para el almacén. El enfoque propuesto se evaluó comparando el rendimiento de los sistemas con y sin la unidad analítica. La red neuronal permite completar la misma cantidad de pedidos con menos robots auxiliares, en menos tiempo y reduce el número de colisiones durante el movimiento de los robots.

En [34] se explora un enfoque para la preparación de pedidos en bodegas considerando robots que pueden tomar los artículos de una estantería (robots recogedores) y robots transportadores que pueden llevar los artículos de una estación de recogida a una

estación de embalaje. Se plantean tres preguntas: i) ¿Qué combinación de robots recolectores y transportadores es la adecuada para obtener un rendimiento superior al de las operaciones de recogida tradicionales basadas en humanos?, ii) ¿Cómo se debería modificar la composición de la flota de robots para optimizar el rendimiento? y iii) ¿Cuál es el impacto del diseño del almacén cuando se despliegan robots móviles coordinados? Concluyen luego de un extenso análisis numérico revela que i) El aumento del número de pasillos transversales disminuye el rendimiento del sistema; ii) las estaciones de empaquetado situadas en el centro mejoran el rendimiento del sistema; y iii) la distancia media desde cada ubicación de recogida hasta la estación de empaquetado y la distancia media entre las ubicaciones de recogida son métricas eficaces para identificar las modificaciones específicas de la flota que probablemente produzcan mejoras en el sistema.

5. Planteamiento del problema

5.1 Ejemplo ilustrativo

Supóngase que un gerente requiere rediseñar las secciones de los productos de una tienda donde se va a considerar el histórico de las transacciones (compras). Para este ejemplo se suponen 24 compras. Los datos de muestra para este ejemplo se presentan en la Tabla 5-1.

Tabla 5-1: Muestra de transacciones (compras).

Id	Transacciones
1	{1 x Leche Deslactosada, 5 x Manzanas Gala Roja, 5 Manzanas Gala Verde, 1 x Paquete de Galletas Saladas, 1 x Jabón Para Ropa}
2	{1 x Leche Entera, 1 x Paquete de Galletas Dulces}
3	{1 x Leche Entera, 8 x Manzanas Gala Roja, 1 x Paquete de Galletas Saladas, 1 x Jabón Para La Piel}
4	{1 x Leche Entera, 15 x Manzanas Gala Verde, 1 x Paquete de Galletas Dulces, 1 x Jabón Para La Piel}
5	{1 x Leche Deslactosada, 1 x Paquete de Galletas Dulces}
6	{1 x Leche Descremada, 1 x Jabón Para Ropa, 1 x Paquete de Galletas Dulces}
7	{1 x Leche Deslactosada, 1 x Paquete de Galletas Dulces, 1 x Jabón Para Ropa}
8	{1 x Leche Deslactosada, 1 x Paquete de Galletas Saladas, 1 x Jabón Para Ropa}
9	{1 x Leche Deslactosada, 1 x Paquete de Galletas Dulces}
10	{1 x Leche Deslactosada, 1 x Paquete de Galletas Saladas}
11	{1 x Leche Deslactosada, 1 x Jabón Para Ropa, 1 x Escoba}
12	{1 x Leche Entera, 2 x Manzanas Gala Verde, 3 x Naranja}
13	{1 x Leche Deslactosada, 1 x Paquete de Galletas Dulces, 1 x Agua Carbonatada}
14	{1 x Leche Entera, 1 x Jamón}
15	{1 x Leche Deslactosada, 1 x Jamón, 1 x Naranja, 1 x Paquete de Galletas Dulces}
16	{1 x Leche Deslactosada, 3 x Agua Carbonatada, 3 x Escoba}
17	{1 x Leche Entera, 1 x Escoba, 1 x Jabón Para Ropa, 1 x Paquete de Galletas

	Dulces}
18	{1 x Leche Deslactosada, 1 x Paquete de Galletas Dulces}
19	{1 x Leche Deslactosada, 4 x Manzanas Gala Roja, 1 x Queso Holandes, 2 x Naranjas}
20	{1 x Leche Deslactosada, 1 x Cepillo De piso, 1 x Agua Mineral, 1 x Paquete de Galletas Dulces}
21	{1 x Leche Entera, 7 x Manzanas Gala Roja, 1 x Queso Holandes, 1 x Jamón, 1 x Paquete de Galletas Saladas}
22	{1 x Leche Deslactosada, 1 x Paquete de Galletas Dulces}
23	{1 x Leche Entera, 1 x Agua Mineral, 1 x Escoba}
24	{1 x Leche Deslactosada, 3 x Escoba}

Los productos de las transacciones presentados en Tabla 5-1 se pueden clasificar en una estructura jerárquica de categorías: secciones, divisiones, grupos, clase y subclases propuesto por el Departamento de Asuntos Económicos y Sociales de las Naciones Unidas [35]. En la Figura 5-1 se presenta la categorización aplicada a nivel de subclases para algunos de los productos de la Tabla 5-1. En la Tabla 5-2 se muestra un ejemplo de estadísticas de las transacciones luego de clasificar los productos en las categorías correspondientes.

Figura 5-1: Árbol de clasificación de productos. Adaptado de [35].

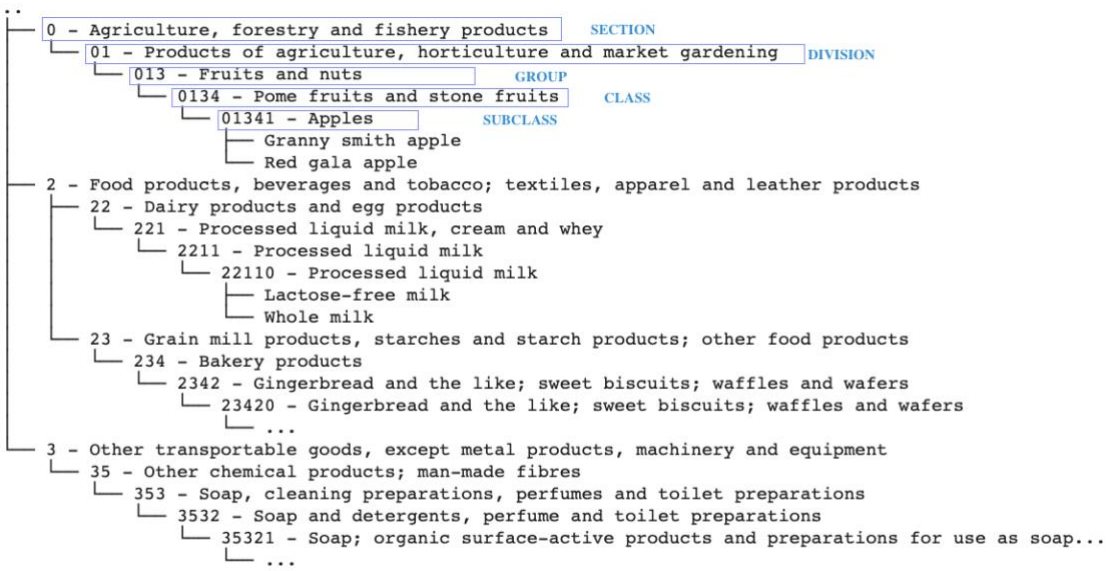


Tabla 5-2: Algunas estadísticas de los productos de la Tabla 5-1 clasificados según el árbol de clasificación de la Figura 5-1**.**

Número de unidades categoría de producto									
ID	***Processed liquid milk***	***Apples***	***Biscuits***	***Soap***	***Oranges***	***Brooms, brushes..***	***Bottled waters...***	***Pig meat …***	***Dairy products***
1	1	10	1	1	0	0	0	0	0
2	1	0	1	0	0	0	0	0	0
3	1	8	1	1	0	0	0	0	0
4	1	15	1	1	0	0	0	0	0
5	1	0	1	0	0	0	0	0	0
6	1	0	1	1	0	0	0	0	0
7	1	0	1	1	0	0	0	0	0
8	1	0	1	1	0	0	0	0	0
9	1	0	1	0	0	0	0	0	0
10	1	0	1	0	0	0	0	0	0
11	1	0	0	1	0	1	0	0	0
12	1	2	0	0	3	0	0	0	0
13	1	0	1	0	0	0	1	0	0
14	1	0	0	0	0	0	0	1	0
15	1	0	1	0	1	0	0	1	0
16	1	0	0	0	0	3	3	0	0
17	1	0	1	1	0	1	0	0	0
18	1	0	1	0	0	0	0	0	0
19	1	4	0	0	2	0	0	0	1
20	1	0	1	0	0	1	1	0	0
21	1	7	1	0	0	0	0	1	1
22	1	0	1	0	0	0	0	0	0
23	1	0	0	0	0	1	1	0	0
24	1	0	0	0	0	3	0	0	0
Número total de unidades de la categoría considerando todas las transacciones	24	46	17	8	6	10	6	3	2
Número de transacciones en que aparece la categoría	24	6	17	8	3	6	4	3	2
%	100%	25%	70.833%	33.333%	12.5%	25%	16.667%	12.5%	8.333%

En este ejemplo se supone que *Processed liquid milk* es adquirido en todas las 24 compras. Según los datos de la Tabla 3 se observa que *Processed liquid milk* se suele adquirir junto con *Biscuits* (en el 70.8% de las transacciones), con *Soap* (en el 33% de las transacciones) y con *Apples* (en el 25% de las transacciones). Esto sugiere que *Processed liquid milk* podría estar en dos o más secciones de la tienda, e.g., en la sección de horneados y en la sección de frutas (pero no en la sección de aseo, por razones sanitarias).

Lo anterior muestra que los conjuntos de productos que se suelen adquirir en conjunto por sí solos no son suficientes para resolver el problema de diseño de una tienda (i.e., ¿cuáles productos se deben colocar contiguos? [36]). Por ejemplo, hay leyes (e.g., no se

permite que los productos químicos estén cerca de los alimentos) o tradiciones (e.g., la ropa interior masculina y femenina no se suelen ubicar en la misma sección) para la ubicación de los productos. Así se permitiera esto podría ocasionar daños a la salud (contaminación de un alimento con un químico) o cuando menos parecer extraño (no se ajusta a "tradiciones establecidas" [36]). Según lo anterior se justifica la especificación de un conjunto de restricciones de productos (o de categorías de productos) para especificar cuáles no deberían estar contiguos, como la leche y el jabón, a pesar de que suelen aparecer juntos en las compras.

Como se indicó en la introducción, para el diseño las secciones de una tienda se puede aplicar la minería de datos [37] y las reglas de asociación para encontrar que productos o cuales categorías de productos deberían ubicarse cerca (en la misma sección). Una regla de asociación [6] es una forma de medir la relación entre los productos y son usadas para descubrir relaciones entre variables en bases de datos de transacciones.

A continuación, se presentan algunos conceptos esenciales de las reglas de asociación. Estas se pueden calcular a partir de las transacciones (compras de los clientes) de un supermercado e indican el comportamiento real del consumidor [1]. A cada uno de los elementos (productos o sus categorías) que forman parte de las transacciones se les conoce como un item y a un conjunto de estos un itemset.

Una regla de asociación tiene dos partes: antecedente o *left hand side* (LHS) y consecuente o *right hand side* (RHS). Por ejemplo, sea la regla $\{ Processed\ liquid\ milk \} \Rightarrow \{Apples\}$. Aquí *Processed liquid milk* es el antecedente y *Apples* es el consecuente. Las reglas de asociación se generan durante la búsqueda de patrones frecuentes [38].

El problema de encontrar reglas de asociación se divide en dos subproblemas: i) encontrar conjuntos de elementos frecuentes ii) encontrar reglas de asociación a partir de estos conjuntos de elementos. Finalmente, se seleccionan las reglas de asociación según métricas como el soporte y la confianza [38], [7]. A continuación se detallan estas funciones:

Si se considera la regla de asociación $A \Rightarrow B$ donde A y B son *itemsets.*

Soporte (*support*): Es el indicador de la frecuencia con la que se produce un elemento en la base de datos. Para una regla $A \Rightarrow B$, el soporte es el porcentaje de transacciones en la base de datos que contienen $A \cup B$ [38].

Para el caso de un *itemset* A el soporte es el número de transacciones que contienen A dividido entre el total de transacciones [39]. Se expresa con las fórmulas:

$$support\ (A) = P(A)$$
$$support\ (A \Rightarrow B) = P(A \cup B)$$

La notación $P(A)$ representa la proporción de veces en el que el *item* A aparece en las transacciones, similarmente $P(A \cup B)$ representa la proporción que A y B coinciden en las transacciones [7]. Como ejemplo si se calcula de la Tabla 4-2 el soporte de { *Processed liquid milk, Apples}* es 25% ya que *Processed liquid milk* y *Apples* aparecen en 6 de 24 transacciones. También se usa la función *support count* que corresponde al número de transacciones que contienen el *itemset*. Por ejemplo $\{Soap\}$ aparece en 8 transacciones.

Confianza (*confidence*): La confianza de las reglas de asociación $A \Rightarrow B$ es la probabilidad de ocurrencia del evento B con base en la ocurrencia del evento A. Se expresa con la fórmula [40]:

$$confidence\ (A \Rightarrow B) = P\ (B \mid A) = \frac{P(A \cup B)}{P(A)}$$

También se puede denotar como [17]$confidence\ (A \Rightarrow B) = \frac{support(A \cup B)}{support(A)}$

Como ejemplo para la regla $\{Apples\} \Rightarrow \{Processed\ liquid\ milk\}$ la confianza es:

$$confidence\ (\{Apples\} \Rightarrow \{Processed\ liquid\ milk\}) = \frac{support(\{Apples\} \cup \{Processed\ liquid\ milk\})}{support(\{Apples\})}$$

Donde:

$$support(\{Processed\ liquid\ milk\ \cup\ Apples\}) = 0.25$$
$$support(\{Apples\}) = 0.25$$

Finalmente:

$$confidence\ (Apples \Rightarrow\ Processed\ liquid\ milk) = \frac{0.25}{0.25} = 1$$

Esto indica que una transacción (compra) que incluya *Apples* tendrá también *Processed liquid milk* con un 100% de probabilidad.

Elevación (*lift*): Indica si el LHS influye positiva o negativamente en el RHS. Esto ayuda a reducir el espacio de búsqueda y, en consecuencia, mejora el rendimiento si es utilizado en la generación de reglas [41]. Por lo tanto, cuanto mayor sea el *lift*, mayor será la probabilidad de que A y B aparezcan juntos. Se expresa con la fórmula [41]:

$$Lift\ (A \Rightarrow B) = \frac{Confidence(A\ \cup\ B)}{P(A) * P(B)} = \frac{Support(A\ \cup\ B)}{Support(B) * Support(A)}$$

Al calcular el *Lift*, se tienen tres posibilidades [41]:

1. Si es mayor que 1, la correlación es positiva.
2. Si es menor que 1, la correlación es negativa.
3. Si es igual a 1, la correlación es independiente.

Como ejemplo para la regla $\{Apples\}\ \Rightarrow \{Processed\ liquid\ milk\}$ el *lift* es:

$$lift\ (\{Apples\} \Rightarrow \{Processed\ liquid\ milk\})$$
$$= \frac{Support(\{Apples\} \cup \{Processed\ liquid\ milk\})}{Support(\{Processed\ liquid\ milk\}) * Support(\{Apples\})}$$

Donde:

$$Support(\{Apples\} \cup \{Processed\ liquid\ milk\}) = 0.25$$
$$Support(\{Processed\ liquid\ milk\}) = 1$$
$$Support(\{Apples\}) = 0.25$$

Finalmente:

$$lift\ (\{Apples\} \Rightarrow \{Processed\ liquid\ milk\}) = \frac{0.25}{1 * 0.25} = 1$$

Como $lift\ (\{Apples\} \Rightarrow \{Processed\ liquid\ milk\}) = 1$ la correlación es independiente, lo que indica que la probabilidad de ocurrencia del LHS y del RHS son independientes.

Existen diferentes algoritmos para generar reglas de asociación, como lo son Apriori [42] y FP-growth [13], [14]. Cualquier algoritmo de reglas de asociación debe encontrar el mismo conjunto de reglas, aunque sus eficiencias computacionales y requisitos de memoria puedan ser diferentes. En esta tesis se usará el algoritmo Apriori.

El algoritmo Apriori es uno de los más usados para encontrar reglas de asociación [1]. Es el algoritmo más famoso y clásico para la minería de patrones frecuentes [38]. El algoritmo tiene dos etapas:

En la primera etapa se encuentran todos los *frequent itemsets* (conjuntos de elementos frecuentes). Los pasos del proceso son los siguientes [40]: Se revisan todas las transacciones para generar el conjunto de elementos candidatos $C1$. Es decir, a partir de las transacciones se generan conjuntos de un elemento (*1-itemsets*) obteniendo $C1$. A continuación, de acuerdo con un umbral de soporte mínimo, se seleccionan los *itemsets* que cumplen, los cuales se denominan *frequent itemsets* $L1$. En la Tabla 5-3 se enseña la primera iteración del algoritmo aplicada a las transacciones de la Tabla 5-1 usando un umbral de soporte mínimo mayor igual que 20%.

Tabla 5-3: Primera iteración Apriori: generación de *frequent itemsets.*

1-itemset $C1$	***Support count***	***Support***	***Frequent itemset $L1$***	***Support***
{Processed liquid milk}	24	1	*{Processed liquid milk}*	1
{Apples}	6	0.25	*{Apples}*	0.25
{Biscuits}	17	0.708	*{Biscuits}*	0.708
{Soap}	8	0.333	*{Soap}*	0.333
{Oranges}	3	0.125	*{Oranges}*	0.125
{Brooms, brushes..}	6	0.25	*{Brooms, brushes..}*	0.25
{Bottled waters...}	4	0.167	*{Bottled waters...}*	0.167
{Pig meat..}	3	0.125	*{Pig meat..}*	0.125
{Dairy products}	2	0.083	*{Dairy products}*	0.083

Todos los *1-itemsets* $C1$ tienen un soporte igual o superior al umbral de soporte establecido, por lo que todos se seleccionan para el conjunto de *frequent itemsets* $L1$. A continuación se generan el conjunto candidato de los posibles *2-itemsets* $C2$ que se pueden obtener a partir de $L1$ y se calcula su soporte. Nuevamente de acuerdo con el umbral de soporte mínimo se seleccionan los elementos para formar los *frequent itemsets* $L2$ como se muestra en la Tabla 5-4.

Tabla 5-4: Segunda iteración Apriori: generación de *frequent itemsets.*

2-itemset $C2$	***Support count***	***Support***	***Frequent itemset*** $L2$	***Support***
{'Processed liquid milk', 'Apples'}	6	0.25	*{'Processed liquid milk', 'Apples'}*	0.25
{'Processed liquid milk', 'Biscuits'}	17	0.708	*{'Processed liquid milk', 'Biscuits'}*	0.708
{'Processed liquid milk', 'Soap'}	8	0.333	*{'Processed liquid milk', 'Soap'}*	0.333
{'Processed liquid milk', 'Oranges'}	3	0.125		
{'Processed liquid milk', 'Brooms, brushes...'}	6	0.25	*{'Processed liquid milk', 'Brooms, brushes...'}*	0.25
{'Processed liquid milk', 'Bottled waters...'}	4	0.167		
{'Processed liquid milk', 'Pig meat...'}	3	0.125		
{'Processed liquid milk', 'Dairy products'}	2	0.083		
{'Apples', 'Biscuits'}	4	0.167		
{'Apples', 'Soap'}	3	0.125		
{'Apples', 'Oranges'}	2	0.083		
{'Apples', 'Brooms, brushes..'}	0	0		
{'Apples', 'Bottled waters...'}	0	0		
{'Apples', 'Pig meat…'}	1	0.042		
{'Apples', 'Dairy products'}	2	0.083		
{'Biscuits', 'Soap'}	7	0.292	*{'Biscuits', 'Soap'}*	0.292
{'Biscuits', 'Oranges'}	1	0.042		
{'Biscuits', 'Brooms, brushes...'}	2	0.083		
{'Biscuits', 'Bottled waters...'}	2	0.083		
{'Biscuits', 'Pig meat…'}	2	0.083		
{'Biscuits', 'Dairy products'}	1	0.042		
{'Soap', 'Oranges'}	0	0		
{'Soap', 'Brooms, brushes...'}	2	0.083		
{'Soap', 'Bottled waters...'}	0	0		
{'Soap', 'Pig meat…'}	0	0		
{'Soap', 'Dairy products'}	0	0		
{'Oranges', 'Brooms', 'brushes...'}	0	0		
{'Oranges', 'Bottled waters...'}	0	0		
{'Oranges', 'Pig meat..'}	1	0.042		
{'Oranges', 'Dairy products'}	1	0.042		
{'Brooms, brushes..', 'Bottled waters...'}	3	0.125		
{'Brooms, brushes...', 'Pig meat...'}	0	0		

{'Brooms, brushes...', 'Dairy products'}	0	0		
{'Bottled waters...', 'Pig meat...'}	0	0		
{'Bottled waters...', 'Dairy products'}	0	0		
{'Pig meat...', 'Dairy products'}	1	0.042		

Se continua similarmente a la iteraciones anteriores obteniendo los 3-itemsets $C3$ y los frequent itemsets $L3$, presentados en la Tabla 5-5, donde no se tienen suficientes *items* para generar 4-itemsets $C4$ deteniendo así el primer paso.

Tabla 5-5: Tercera iteración Apriori: generación de *frequent itemsets.*

3-itemset $C3$	***Support count***	***Support***	***Frequent itemset $L3$***	***Support***
{'Processed liquid milk', 'Apples', 'Biscuits'}	4	0.167		
{'Processed liquid milk', 'Apples', 'Soap'}	3	0.125		
{'Processed liquid milk', 'Apples', 'Brooms, brushes...'}	0	0		
{'Processed liquid milk', 'Biscuits', 'Soap'}	7	0.292	*{'Processed liquid milk', 'Biscuits', 'Soap'}*	0.292
{'Processed liquid milk', 'Biscuits', 'Brooms, brushes...'}	2	0.083		
{'Processed liquid milk', 'Soap', 'Brooms, brushes...'}	2	0.083		
{'Apples', 'Biscuits', 'Soap'}	3	0.125		
{'Apples', 'Biscuits', 'Brooms, brushes...'}	0	0		
{'Apples', 'Soap', 'Brooms, brushes...'}	0	0		
{'Biscuits', 'Soap', 'Brooms, brushes...'}	1	0.042		

El siguiente paso es generar reglas de asociación a partir de cada uno de los *frequent itemsets*, para esto se aplica el siguiente algoritmo [43]:

Para cada *frequent itemset* L, obtener todos los posibles subsets no vacíos de L

Para cada subset s de L, crear la regla $s \Rightarrow (L - s)$

Descartar todas las reglas que no superen un mínimo de confianza

Supóngase un mínimo de confianza de 25%, al calcular las reglas de asociación a partir de los *frequent itemsets* de las Tablas 5-3, 5-4 y 5-5 se muestran los resultados en la Tabla 5-6.

Tabla 5-6: Reglas de asociación generadas.

Antecedent	***Consequent***	***Suppport***	***Confidence***
{Brooms - brushes}	*{Processed liquid milk}*	0.25	1
{Processed liquid milk}	*{Brooms - brushes}*	0.25	0.25
{Apples}	*{Processed liquid milk}*	0.25	1
{Processed liquid milk}	*{Apples}*	0.25	0.25
{Soap}	*{Biscuits}*	0.292	0.875
{Biscuits}	*{Soap}*	0.292	0.412
{Soap}	*{Processed liquid milk}*	0.333	1
{Processed liquid milk}	*{Soap}*	0.333	0.333
{Biscuits}	*{Processed liquid milk}*	0.708	1
{Processed liquid milk}	*{Biscuits}*	0.708	0.708
{Biscuits,Soap}	*{Processed liquid milk}*	0.292	1
{Processed liquid milk,Soap}	*{Biscuits}*	0.292	0.875
{Biscuits,Processed liquid milk}	*{Soap}*	0.292	0.412

En la Tabla 5-6 si se consideran las reglas de *Biscuits* ⇒ *Processsed liquid milk* (con un soporte: 70.8%) y *Apples* ⇒ *Proccessed liquid milk* (con un soporte 25%) y se supone un umbral mínimo de soporte del 30%, entonces la regla *Apples* ⇒ *Proccessed liquid milk* no se consideraría significativa; sin embargo, el producto *Apples* es el más vendido (46 unidades). Este aspecto es ignorado en la minería de reglas de asociación clásica y conduce a una pérdida de información [15].

6. Algoritmo

En esta tesis para resolver el problema de diseño expuesto se propone un algoritmo que a partir del histórico de las compras de productos en una tienda pretende generar uno o más diseños para esta. En cada diseño un mismo producto podría aparecer en varias secciones de la tienda, e.g., la leche en la sección de horneados y en la sección de frutas. A continuación se describen las entradas del algoritmo.
Las definiciones de los conjuntos C, P y T son tomados a partir de [44].

6.1 Entradas

- Sea $C = \{c_1,\ c_2, \ldots, c_n\}$ el conjunto de nodos del árbol donde c_i representa la categoría de productos. $c_i = (c_{id}, c_{name}, c_{parent})$, donde $c_{id} \in$ I+ es el identificador único de cada categoría, c_{name} (una cadena de caracteres) es el nombre único de la categoría y c_{parent} es el identificador (c_{id}) de un nodo $c_J \in C$ ($c_i \neq c_j$). i.e., c_j es la categoría padre de c_i (si c_i es la raíz entonces c_i no tiene padre).

- Sea $P = \{p_1,\ p_2, \ldots, p_q\}$ donde P_i representa un producto. $p_i = (p_{id},\ p_{name},\ p_{parent})$, donde $p_{id} \in$ I$^+$ es el identificador único de cada producto, p_{name} es el nombre de cada producto (también único), p_{parent} es el identificador único (c_{id}) del nodo $c_i \in$ C. i.e., la categoría (padre) a la que p_i pertenece.

- Sea $T = \{t_1,\ t_2..t_n\}$ es el conjunto de transacciones donde T_i representa una transacción. $t = (t_{id},\ D)$, donde $t_{id} \in$ I$^+$ es el identificador único de cada transacción y $D = \{d_1,\ d_2..d_m\}$ es un conjunto de detalles. $d_j = (d_{product},\ d_w)$, donde $d_{product}$ es el identificador único (P_{id}) del producto P_i en P, y $d_w \in$ I$^+$ es el número de unidades compradas de un producto P_i en d_j. No puede haber en D dos detalles $d_j,\ d_i, i \neq j$, con el mismo $d_{product}$, i.e, No puede haber en D dos

detalles en la misma transacción con el mismo producto. Ejemplo: $T = \{t_1,\ t_2,\ t_3,\ t_4\}$, donde:

- $t_1 = (1,\ \{(4,2),(5,5),(6,2)\})$
- $t_2 = (2,\ \{(8,3),(5,5),(7,3)\})$
- $t_3 = (3,\ \{(2,2)\})$

- Sean $minssupp, minconf$ y $minlift$ constantes $\in R^+$ que indican el umbral de *support*, *confidence* y *lift* correspondientemente que debe de cumplir una regla de una regla de asociación R para ser considerada significativa.
  - $Support(R) \geq minsupp$
  - $Confidence(R) \geq minconf$
  - $Lift(R) \geq minlift$

- Sea $RE = \{R_1,\ R_2 \dots R_n\}$ el conjunto de restricciones aplicadas al árbol de clasificación de productos donde R_i representa una restricción. $R_i = (R_{id}, R_{product\ category\ 1}, R_{product\ category\ 2})$, donde $R_{id} \in$ I$^+$ es el identificador único de cada restricción, $R_{product\ category\ 1}$ es el identificador único (c_{id}) del nodo $c_i \in C$ para la categoría de productos que no puede estar cercana a la categoría $R_{product\ category\ 2}$ que es el identificador único (c_{id}) del nodo $c_i \in C$. En esta tesis solamente se consideran restricciones de este tipo. Se tienen además las siguientes consideraciones:

- **Restricciones equivalentes**: Dada una restricción R_a= $(R_{ida}, R_{product\ category\ a1}, R_{product\ category\ a2})$ y la restricción R_b= $(R_{idb}, R_{product\ category\ a2}, R_{product\ category\ a1})$, entonces solo una de las dos debe de ser considerada porque son equivalentes. Por ejemplo, las restricciones: R_a= $(1, Dairy\ products\ and\ egg\ products\ , Pome\ fruits\ and\ stone\ fruits)$ y R_b= $(2, Pome\ fruits\ and\ stone\ fruits, Dairy\ products\ and\ egg\ products\ )$ son equivalentes.

- **Restricción de categorías diferentes**: Dada una restricción R_a= $(R_{ida}, R_{product\ category\ a1}, R_{product\ category\ a2})$ entonces las dos categorías de

productos deben de ser diferentes, i.e., $product\ category\ a1 \neq product\ category\ a2$. Por ejemplo, la restricción: R_a= $(1, Dairy\ products\ and\ egg\ products\ , Dairy\ products\ and\ egg\ products)$ no está permitida.

- **Restricción dentro de un mismo ramal**: Dada una restricción R_a= $(R_{ida}, R_{product\ category\ a}, R_{product\ category\ b})$ entonces $R_{product\ category\ b}$ no puede ser descendiente de $R_{product\ category\ a}$ o $R_{product\ category\ a}$ no puede ser descendiente de $R_{product\ category\ b}$. Por ejemplo, la restricción R_a= $(1, Dairy\ products\ and\ egg\ products\ , Processed\ liquid\ milk)$ o $(1\ , Processed\ liquid\ milk, Dairy\ products\ and\ egg\ products)$ no está permitida.

- **Restricción raíz**: El nodo raíz, el cual se denomina *All products*, no puede ser parte de ninguna restricción. Por ejemplo, la restricción: R_a= $(1, All\ products\ , processed\ liquid\ milk)$ no está permitida.

En esta tesis solo se consideran las restricciones presentadas. Así, no se van a considerar restricciones que involucren por ejemplo tres categorías, como: C_1 puede estar con C_2 y C_1 puede estar con C_3 pero C_1 no puede simultáneamente con C_2 y C_3.

6.2 Paso 1: Reglas de asociación y la cantidad

A partir de T (conjunto de transacciones) se calculan las reglas de asociación mediante el algoritmo Apriori. A continuación aplicando la metodología propuesta en [17], el conjunto de reglas de asociación generadas se somete a restricciones de ponderación ($Wgain$) y utilidad ($Ugain$), y para cada regla de asociación, se calcula una puntuación ponderada de utilidad combinada ($UWScore$) [17]. Finalmente, se determina un subconjunto de reglas de asociación a considerar en función de la puntuación $UWScore$.

Los pasos propuestos en [17] son:

1. Cálculo de reglas de asociación usando el algoritmo Apriori.
2. Cálculo de la medida *Wgain*.
3. Cálculo de la medida *Ugain*.
4. Cálculo de *UWscore* a partir de *Wgain* y *Ugain*.

5. Selección de reglas de asociación significativas a partir del *UWscore*.

A continuación, se describen los pasos.

6.2.1 Paso 1.1: Minería de reglas de asociación mediante el algoritmo Apriori.

Las reglas de asociación se generan a partir del conjunto de transacciones T. Cada transacción T tiene un número m de detalles $D = \{d_1,\ d_2, \ldots, d_m\}$ donde $d_j = (d_{product},\ d_w)$, d_w es la cantidad de unidades adquirida del producto $d_{product}$.

El algoritmo Apriori genera un número k de reglas de asociación $R = \{R_1,\ R_2 \ldots, R_k\}$. El conjunto de reglas de asociación R es la entrada al siguiente paso: el cálculo de la ponderación y la utilidad. Las medidas $Wgain$ (ponderación) y $Ugain$ (utilidad) se calculan para cada atributo (ítem) de las k reglas de asociación de R.

Por ejemplo, sea una regla de asociación R_i de la forma, $(A, B) \Rightarrow C$, donde, A, B y C son los atributos en la regla R_i, las medidas $Ugain$ y $Wgain$ se calculan para A, B y C y luego se calcula $UWscore$ para cada regla R_i. Inicialmente, las k reglas de asociación generadas se clasifican en orden descendente según su nivel de confianza. La lista ordenada de reglas es $S = \{R_1',\ R_2', \ldots, R_k'\}$.

Al calcular las reglas de asociación sobre las transacciones de la Tabla 5-1 mediante el algoritmo Apriori con un soporte mínimo de 0.05 y una confianza mínima de 0.5, ordenadas por nivel de confianza se obtiene, e.g., las reglas mostradas en la Tabla 6-1.

Tabla 6-1: Reglas de asociación ordenadas descendentemente por *confidence* generadas de las transacciones de la Tabla 5-1.

# *Rule*	*Antecedent*	*Consequent*	*Support*	*Confi-dence*	*Lift*
1	*{Biscuits}*	*{Processed liquid milk}*	0.708	1	1
2	*{Soap}*	*{Processed liquid milk}*	0.333	1	1
3	*{Biscuits,Soap}*	*{Processed liquid milk}*	0.292	1	1
4	*{Brooms - brushes}*	*{Processed liquid milk}*	0.25	1	1
5	*{Apples}*	*{Processed liquid milk}*	0.25	1	1
6	*{Bottled waters}*	*{Processed liquid milk}*	0.167	1	1
7	*{Apples,Biscuits}*	*{Processed liquid milk}*	0.167	1	1
8	*{Apples,Soap}*	*{Biscuits}*	0.125	1	1.412
9	*{Apples,Processed liquid milk,Soap}*	*{Biscuits}*	0.125	1	1.412
10	*{Oranges}*	*{Processed liquid milk}*	0.125	1	1

11	*{Pig meat}*	*{Processed liquid milk}*	0.125	1	1
12	*{Bottled waters,Brooms - brushes}*	*{Processed liquid milk}*	0.125	1	1
13	*{Apples,Soap}*	*{Processed liquid milk}*	0.125	1	1
14	*{Apples,Biscuits,Soap}*	*{Processed liquid milk}*	0.125	1	1
15	*{Dairy products}*	*{Apples}*	0.083	1	4
16	*{Dairy products,Processed liquid milk}*	*{Apples}*	0.083	1	4
17	*{Dairy products}*	*{Processed liquid milk}*	0.083	1	1
18	*{Apples,Dairy products}*	*{Processed liquid milk}*	0.083	1	1
19	*{Apples,Oranges}*	*{Processed liquid milk}*	0.083	1	1
20	*{Biscuits,Pig meat}*	*{Processed liquid milk}*	0.083	1	1
21	*{Biscuits,Bottled waters}*	*{Processed liquid milk}*	0.083	1	1
22	*{Brooms - brushes,Soap}*	*{Processed liquid milk}*	0.083	1	1
23	*{Biscuits,Brooms - brushes}*	*{Processed liquid milk}*	0.083	1	1
24	*{Soap}*	*{Biscuits}*	0.292	0.875	1.235
25	*{Processed liquid milk,Soap}*	*{Biscuits}*	0.292	0.875	1.235
26	*{Bottled waters}*	*{Brooms - brushes}*	0.125	0.75	3
27	*{Bottled waters,Processed liquid milk}*	*{Brooms - brushes}*	0.125	0.75	3
28	*{Apples,Biscuits}*	*{Soap}*	0.125	0.75	2.25
30	*{Processed liquid milk}*	*{Biscuits}*	0.708	0.708	1
31	*{Apples}*	*{Biscuits}*	0.167	0.667	0.941
32	*{Apples,Processed liquid milk}*	*{Biscuits}*	0.167	0.667	0.941
33	*{Oranges}*	*{Apples}*	0.083	0.667	2.667
34	*{Oranges,Processed liquid milk}*	*{Apples}*	0.083	0.667	2.667
35	*{Pig meat}*	*{Biscuits}*	0.083	0.667	0.941
36	*{Pig meat,Processed liquid milk}*	*{Biscuits}*	0.083	0.667	0.941
37	*{Brooms - brushes}*	*{Bottled waters}*	0.125	0.5	3
38	*{Brooms - brushes,Processed liquid milk}*	*{Bottled waters}*	0.125	0.5	3
39	*{Apples}*	*{Soap}*	0.125	0.5	1.5
40	*{Apples,Processed liquid milk}*	*{Soap}*	0.125	0.5	1.5
41	*{Bottled waters}*	*{Biscuits}*	0.083	0.5	0.706
42	*{Bottled waters,Processed liquid milk}*	*{Biscuits}*	0.083	0.5	0.706

6.2.2 Paso 1.2: Cálculo de la medida Wgain

De la lista ordenada de reglas obtenida en el paso anterior S, se selecciona la primera regla $R_1{}'$. Para cada uno de los atributos de $R_1{}'$ se calcula la medida $Wgain$, la cual es la suma de los pesos de los elementos d_w de los detalles para cada transacción del conjunto T .

$$Wgain_j = \sum_{i=1}^{|T|} d_{w(i,j)}$$

Donde $d_{w(i,j)}$ representa el peso (cantidad) del producto j en la transacción i y $|T|$ es el número de transacciones en T (corresponde a n). En la fórmula anterior se calcula la cantidad que se adquiere un producto en todas las transacciones, con el fin de comparar que tan significativa es en comparación con los demás productos.

Al calcular $Wgain$ para las primeras cinco reglas de la Tabla 6-1 se obtiene la Tabla 6-2

Tabla 6-2: *Wgain* de las primeras 5 reglas de asociación de la Tabla 6-1.

# ***Rule***	***Rule***	***Attribute***	***W-gain***
1	*{Biscuits} ⇒ {Processed liquid milk}*	*Biscuits*	17
		Processed liquid milk	24
2	*{Soap} ⇒ {Processed liquid milk}*	*Soap*	8
		Processed liquid milk	24
3	*{Biscuits,Soap} ⇒ {Processed liquid milk}*	*Biscuits*	17
		Soap	8
		Processed liquid milk	24
4	*{Brooms - brushes} ⇒ {Processed liquid milk}*	*Brooms - brushes*	10
		Processed liquid milk	24
5	*{Apples} ⇒ {Processed liquid milk}*	*Apples*	46
		Processed liquid milk	24

6.2.3 Paso 1.3: Cálculo de la medida *Ugain*

Similarmente, para calcular la medida $Ugain$, se selecciona la primera regla $R_1{}'$ de la lista ordenada. Para cada uno de los atributos de $R_1{}'$ se calcula la medida $Ugain$, el cual se basa en el $Ufactor$ y en el valor de utilidad U_j de cada producto P.

El conjunto de utilidades es:

$$U = \{U_1, U_2 \ \dots \ U_q\} \ donde \ U > 0$$

El factor de utilidad $Ufactor$ es una constante determinada por la suma de las utilidades de todos los productos:

$$Ufactor = \frac{1}{\sum_{i=1}^{q} U_i}$$

Donde $Ufactor > 0$ y q corresponde a la cantidad de productos únicos.

Aquí se considera la utilidad de cada producto como una constante c, i.e., todos los productos tienen la misma utilidad. Para ejemplificar, si se supone una utilidad U = 1 para todos los productos se obtiene:

$$Ufactor = \frac{1}{\sum_{i=1}^{q} 1}$$

Que es equivalente a:

$$Ufactor = \frac{1}{q}$$

Se define $Ugain_j$ para obtener la relación de utilidad de cada producto con relación a la suma total de utilidades. Con el fin de saber qué tan significativo es un producto en comparación con los demás con respecto a su utilidad.

Es decir:

$$Ugain_j = U_j * Ufactor$$

Como $Ufactor$ es una constante entonces:

$$Ugain_j = c' Donde\ Ugain_j, c' > 0$$

Por ejemplo, al calcular $Ugain$ para la primera regla de la Tabla 6-1, con $q = 9$ se obtiene:

$$Ugain_{R_1} = \frac{1}{\sum_{i=1}^{9} 1} = \frac{1}{9} = 0.111$$

Al calcular $Ugain$ para las primeras cinco reglas de la Tabla 6-1 se obtiene la Tabla 6-3. Nótese que se obtiene una constante 0.111 de $Ugain$ para todos los atributos de las reglas.

Tabla 6-3: *Ugain* de las primeras 5 reglas de asociación de la Tabla 6-1.

# Rule	***Rule***	***Attribute***	***W-gain***	***U-gain***
1	*{Biscuits} ⇒ {Processed liquid milk}*	*Biscuits*	17	0.111
		Processed liquid milk	24	0.111

2	*{Soap} ⇒ {Processed liquid milk}*	*Soap*	8	0.111
		Processed liquid milk	24	0.111
3	*{Biscuits,Soap} ⇒ {Processed liquid milk}*	*Biscuits*	17	0.111
		Soap	8	0.111
		Processed liquid milk	24	0.111
4	*{Brooms - brushes} ⇒ {Processed liquid milk}*	*Brooms - brushes*	10	0.111
		Processed liquid milk	24	0.111
5	*{Apples} ⇒ {Processed liquid milk}*	*Apples*	46	0.111
		Processed liquid milk	24	0.111

6.2.4 Paso 1.4: Cálculo de *UWscore* a partir de *Wgain* y *Ugain*.

A partir de las medidas de $Wgain$ y $Ugain$ calculadas para los atributos individuales de una regla de asociación R, se calcula un único valor consolidado denominado $UWScore$ para R, así:

$$UWScore_R = \frac{\sum_{j=1}^{|R|}(Wgain)_j * (Ugain)_j}{|R|}$$

Donde $|R|$ representa el número de atributos en la regla de asociación.

Como $Ugain_j$ es igual para todos los productos la fórmula se simplifica así:

$$UWScore_R = c' * \frac{\sum_{j=1}^{|R|}(Wgain)_j}{|R|}$$

Al multiplicar la sumatoria por la constante c' se puede omitir ya que el efecto es el mismo para todas.

$$UWScore_R = \frac{\sum_{j=1}^{|R|}(Wgain)_j}{|R|}$$

Esta fórmula anterior se puede interpretar como la utilidad combinada basada en la ponderación de la cantidad y de la utilidad.

Por ejemplo, para la regla R_5: *{Apples} ⇒ {Processed liquid milk}* se obtuvo *UW-score* = 35 porque:

$$UWScore_{R_5} = \frac{\sum_{j=1}^{|R_5|}(Wgain)_j}{|R_5|} = \frac{\sum_{j=1}^{2}(Wgain)_j}{2} = \frac{46+24}{2} = 35$$

Los procesos para obtener $Wgain$, $Ugain$ y $UWScore$ se repiten para las demás reglas de asociación de S. Ahora cada una las reglas de asociación posee un $UWScore$, con este valor se ordenan las reglas en orden descendente para obtener S'.

En la Tabla 6-4 se muestran las reglas con los cinco mejores de $UWScore$ obtenidos sobre las reglas de la Tabla 6-1.

Tabla 6-4: *UWScore* de las cinco mejores reglas de asociación de la Tabla 6-1.

# Rule	***Rule***	***Attribute***	***W-gain***	***|R|***	***UW-score***	***Support***	***Confidence***
5	*{Apples} ⇒ {Processed liquid milk}*	Apples	46	2	35	0.25	1
		Processed liquid milk	24				
31	*{Apples} ⇒ {Biscuits}*	Apples	46	2	31.5	0.167	0.667
		Biscuits	17				
7	*{Apples,Biscuits} ⇒ {Processed liquid milk}*	Apples	46	3	29	0.167	1
		Biscuits	17				
		Processed liquid milk	24				
32	*{Apples,Processed liquid milk} ⇒ {Biscuits}*	Apples	46	3	29	0.167	0.667
		Processed liquid milk	24				
		Biscuits	17				
39	*{Apples} ⇒ {Soap}*	Apples	46	2	27	0.125	0.5
		Soap	8				

6.2.5 Paso 1.5: Selección de reglas de asociación significativas a partir de *UWscore*.

Por último, a partir de la lista ordenada S' que corresponde al conjunto de reglas de asociación ponderadas, se seleccionan las reglas que satisfacen un umbral mínimo para $UWScore$. En el ejemplo si se continua con las demás reglas de la Tabla 6-1 y se selecciona un umbral de $UWScore \geq 14$ se obtiene la Tabla 6-5.

Tabla 6-5: Reglas de asociación de Tabla 6-1 que cumplen con el umbral de $UWScore \geq 14$.

# Rule	Rule	Attribute	W-gain	|R|	UW-score	Support	Confidence
5	*{Apples} ⇒ {Processed liquid milk}*	*Apples*	46	2	35	0.25	1
		Processed liquid milk	24				
31	*{Apples} ⇒ {Biscuits}*	*Apples*	46	2	31.5	0.167	0.667
		Biscuits	17				
7	*{Apples,Biscuits} ⇒ {Processed liquid milk}*	*Apples*	46	3	29	0.167	1
		Biscuits	17				
		Processed liquid milk	24				
32	*{Apples,Processed liquid milk} ⇒ {Biscuits}*	*Apples*	46	3	29	0.167	0.667
		Processed liquid milk	24				
		Biscuits	17				
39	*{Apples} ⇒ {Soap}*	*Apples*	46	2	27	0.125	0.5
		Soap	8				
13	*{Apples, Soap} ⇒ {Processed liquid milk}*	*Apples*	46	3	26	0.125	1
		Soap	8				
		Processed liquid milk	24				
33	*{Oranges} ⇒ {Apples}*	*Oranges*	6	2	26	0.083	0.667
		Apples	46				
40	*{Apples, Processed liquid milk} ⇒ {Soap}*	*Apples*	46	3	26	0.125	0.5
		Processed liquid milk	24				
		Soap	8				
19	*{Apples,Oranges} ⇒ {Processed liquid milk}*	*Apples*	46	3	25.333	0.083	1
		Oranges	6				
		Processed liquid milk	24				
34	*{Oranges,Processed liquid milk} ⇒ {Apples}*	*Oranges*	6	3	25.333	0.083	0.667
		Processed liquid milk	24				
		Apples	46				
15	*{Dairy products} ⇒ {Apples}*	*Dairy products*	2	2	24	0.083	1
		Apples	46				
	{Dairy products,Processed liquid milk} ⇒ {Apples}	*Processed liquid milk*	24	3	24	0.083	1
		Apples	46				
18	*{Apples,Dairy products} ⇒ {Processed liquid milk}*	*Apples*	46	3	24	0.083	1
		Dairy products	2				
		Processed liquid milk	24				
9	*{Apples,Processed liquid milk,Soap} ⇒ {Biscuits}*	*Apples*	46	4	23.75	0.125	1
		Processed liquid milk	24				
		Soap	8				
		Biscuits	17				
14	*{Apples,Biscuits,Soap} ⇒ {Processed liquid milk}*	*Apples*	46	4	23.75	0.125	1
		Biscuits	17				

		Soap	8				
		Processed liquid milk	24				
29	*{Apples,Biscuits,Processed liquid milk} ⇒ {Soap}*	*Apples*	46	4	23.75	0.125	0.75
		Biscuits	17				
		Processed liquid milk	24				
		Soap	8				
8	*{Apples,Soap} ⇒ {Biscuits}*	*Apples*	46	3	23.667	0.125	1
		Soap	8				
		Biscuits	17				
28	*{Apples,Biscuits} ⇒ {Soap}*	*Apples*	46	3	23.667	0.125	0.75
		Biscuits	17				
		Soap	8				
1	*{Biscuits} ⇒ {Processed liquid milk}*	*Biscuits*	17	2	20.5	0.708	1
		Processed liquid milk	24				
30	*{Processed liquid milk} ⇒ {Biscuits}*	*Processed liquid milk*	24	2	20.5	0.708	0.708
		Biscuits	17				
4	*{Brooms - brushes} ⇒ {Processed liquid milk}*	*Brooms - brushes*	10	2	17	0.25	1
		Processed liquid milk	24				
23	*{Biscuits,Brooms - brushes} ⇒ {Processed liquid milk}*	*Biscuits*	17	3	17	0.083	1
		Brooms - brushes	10				
		Processed liquid milk	24				
3	*{Biscuits,Soap} ⇒ {Processed liquid milk}*	*Biscuits*	17	3	16.333	0.292	1
		Soap	8				
		Processed liquid milk	24				
25	*{Processed liquid milk, Soap} ⇒ {Biscuits}*	*Processed liquid milk*	24	3	16.333	0.292	0.875
		Soap	8				
		Biscuits	17				
2	*{Soap} ⇒ {Processed liquid milk}*	*Soap*	8	2	16	0.333	1
		Processed liquid milk	24				
21	*{Biscuits,Bottled waters} ⇒ {Processed liquid milk}*	*Biscuits*	17	3	15.667	0.083	1
		Bottled waters	6				
		Processed liquid milk	24				
42	*{Bottled waters,Processed liquid milk} ⇒ {Biscuits}*	*Bottled waters*	6	3	15.667	0.083	0.5
		Processed liquid milk	24				
		Biscuits	17				
6	*{Bottled waters} ⇒ {Processed liquid milk}*	*Bottled waters*	6	2	15	0.167	1
		Processed liquid milk	24				
10	*{Oranges} ⇒ {Processed liquid milk}*	*Oranges*	6	2	15	0.125	1
		Processed liquid	24				

		milk					
20	*{Biscuits, Pig meat} ⇒ {Processed liquid milk}*	*Biscuits*	17	3	14.667	0.083	1
		Pig meat	3				
		Processed liquid milk	24				
36	*{Pig meat, Processed liquid milk} ⇒ {Biscuits}*	*Pig meat*	3	3	14.667	0.083	0.667
		Processed liquid milk	24				
		Biscuits	17				
22	*{Brooms - brushes, Soap} ⇒ {Processed liquid milk}*	*Brooms - brushes*	10	3	14	0.083	1
		Soap	8				
		Processed liquid milk	24				

6.3 Paso 2: Selección de reglas con *support*, confidence y *lift* significativos

Con las reglas de asociación S' obtenidas del paso anterior se descartan las que no cumplan el soporte mínimo ($suppmin$), la confianza mínima ($minconf$) y el lift mínimo ($minlift$).

Las reglas de $S' = \{R'_{1,} .. R'_{k,}\}$ que cumplan con:

$$Support(R'_i) \geq minsupp$$
$$Confidence(R'_i) \geq minconf$$
$$lift(R'_i) \geq minlift$$

Son seleccionadas y se denominaran S''.

En el ejemplo de la Tabla 6-5 si se toma un $minsupp = 0.1$, $minconf = 0.5$ y $minlift = 0.8$ se obtiene la Tabla 6-6 la cual contiene 20 reglas.

Tabla 6-6: Reglas de asociación resultantes de la Tabla 6-5 que cumplen con los umbrales $minsupp = 0.1$, $minconf = 0.5$ y $minlift = 0.8$.

# Rule	***Rule***	***Attribute***	***W-gain***	***|R|***	***UW-score***	***Support***	***Confidence***	***Lift***
5	*{Apples} ⇒ {Processed liquid milk}*	*Apples*	46	2	35	0.25	1	1
		Processed liquid milk	24					

31	*{Apples} ⇒ {Biscuits}*	*Apples*	46	2	31.5	0.167	0.667	0.941
		Biscuits	17					
7	*{Apples,Biscuits} ⇒ {Processed liquid milk}*	*Apples*	46	3	29	0.167	1	1
		Biscuits	17					
		Processed liquid milk	24					
32	*{Apples,Processed liquid milk} ⇒ {Biscuits}*	*Apples*	46	3	29	0.167	0.667	0.941
		Processed liquid milk	24					
		Biscuits	17					
39	*{Apples} ⇒ {Soap}*	*Apples*	46	2	27	0.125	0.5	1.5
		Soap	8					
13	*{Apples,Soap} ⇒ {Processed liquid milk}*	*Apples*	46	3	26	0.125	1	1
		Soap	8					
		Processed liquid milk	24					
40	*{Apples,Processed liquid milk} ⇒ {Soap}*	*Apples*	46	3	26	0.125	0.5	1.5
		Processed liquid milk	24					
		Soap	8					
9	*{Apples,Processed liquid milk,Soap} ⇒ {Biscuits}*	*Apples*	46	4	23.75	0.125	1	1.412
		Processed liquid milk	24					
		Soap	8					
		Biscuits	17					
14	*{Apples,Biscuits,Soap} ⇒ {Processed liquid milk}*	*Apples*	46	4	23.75	0.125	1	1
		Biscuits	17					
		Soap	8					
		Processed liquid milk	24					
29	*{Apples,Biscuits,Processed liquid milk} ⇒ {Soap}*	*Apples*	46	4	23.75	0.125	0.75	2.25
		Biscuits	17					
		Processed liquid milk	24					
		Soap	8					
8	*{Apples,Soap} ⇒ {Biscuits}*	*Apples*	46	3	23.67	0.125	1	1.412
		Soap	8					
		Biscuits	17					
28	*{Apples,Biscuits} ⇒ {Soap}*	*Apples*	46	3	23.67	0.125	0.75	2.25
		Biscuits	17					
		Soap	8					
1	*{Biscuits} ⇒ {Processed liquid milk}*	*Biscuits*	17	2	20.5	0.708	1	1
		Processed liquid milk	24					
30	*{Processed liquid milk} ⇒ {Biscuits}*	*Processed liquid milk*	24	2	20.5	0.708	0.708	1
		Biscuits	17					
4	*{Brooms - brushes} ⇒ {Processed liquid milk}*	*Brooms - brushes*	10	2	17	0.25	1	1
		Processed liquid milk	24					
3	*{Biscuits,Soap} ⇒ {Processed liquid milk}*	*Biscuits*	17	3	16.33	0.292	1	1
		Soap	8					
		Processed liquid milk	24					
25	*{Processed liquid milk,Soap} ⇒*	*Processed liquid milk*	24	3	16.33	0.292	0.875	1.235

	{Biscuits}	*Soap*	8					
		Biscuits	17					
2	*{Soap} ⇒ {Processed liquid milk}*	*Soap*	8	2	16	0.333	1	1
		Processed liquid milk	24					
6	*{Bottled waters} ⇒ {Processed liquid milk}*	*Bottled waters*	6	2	15	0.167	1	1
		Processed liquid milk	24					
10	*{Oranges} ⇒ {Processed liquid milk}*	*Oranges*	6	2	15	0.125	1	1
		Processed liquid milk	24					

6.4 Paso 3: Selección de reglas que cumplen las restricciones

Supóngase que existen restricciones para la disposición de los productos, e.g., los productos del grupo Fruits and nuts no pueden estar cerca de los productos de la categoría “Soap and detergents…”.

En este paso a partir de las reglas resultantes del paso anterior se descartan las reglas en cuyos atributos o sus descendientes incumplan alguna de las restricciones del conjunto de restricciones $RE = \{RE_1,\ RE_2 \ldots RE_n\}$. Para esto se calcula para cada restricción RE_i todos los descendientes de sus categorías, así:

$$DE_i = \{(R_{product\ category\ 1}, descendientes(R_{product\ category\ 1})),$$
$$(R_{product\ category\ 2}, descendientes(R_{product\ category\ 2}))\}$$

Por ejemplo, si se tiene el conjunto de restricciones $RE = \{RE_1,\ RE_2,\ RE_3,\ RE_4,\ RE_5,\ RE_6,\ RE_7\}$, donde:

$RE_1 = (1, "Processed\ liquid\ milk\ cream\ and\ whey", "Soap\ and\ detergents,\ perfume\ and\ toilet\ preparations")$

$RE_2 = (2, "Apples", "Soap\ and\ detergents, perfume\ and\ toilet\ preparations")$

$RE_3 = (3, "Biscuits", "Soap\ and\ detergents, perfume\ and\ toilet\ preparations")$

$RE_4 = (4, "Oranges", "Soap\ and\ detergents, perfume\ and\ toilet\ preparations")$

$RE_5 = (5, "Bottled\ waters", "Soap\ and\ detergents, perfume\ and\ toilet\ preparations")$

$RE_6 = (6, "Pig\ meat", "Soap\ and\ detergents, perfume\ and\ toilet\ preparations")$

$RE_7 = (7, "Dairy\ Products", "Soap\ and\ detergents, perfume\ and\ toilet\ preparations")$

Al calcular los descendientes DE_1 se obtiene:

$$DE_1 = \left\{ \begin{array}{c} (\text{"}Processed\ liquid\ milk\ cream\ and\ whey\text{"}, \text{"}\ Processed\ liquid\ milk\text{"}, \\ \text{"}\ Processed\ liquid\ milk\text{"}, \text{"}\ Lactose - free\ milk\text{"}, \text{"}\ Whole\ milk\text{"}), \\ (\text{"}Soap\ and\ detergents\text{"}, \text{"}Perfume\ and\ toilet\ preparations\text{"}, \\ \text{"}Soap; organic\ surface - active\ products\ and\ preparations\ for\ use\ as\ soap \ldots\text{"}, \ldots) \end{array} \right\}$$

Nótese que para los descendientes de *Processed liquid milk cream and whey* se repite *Processed liquid milk* en dos ocasiones, esto ocurre porque según la clasificación propuesta por el Departamento de Asuntos Económicos y Sociales de las Naciones Unidas [35] la clase y las subclases llevan el mismo nombre para este caso.

Similarmente, se calculan $DE_{2,}$ DE_3, DE_4, DE_5, DE_6 y DE_7

Al aplicar el conjunto de restricciones RE sobre las reglas de la Tabla 6-6 se obtiene la Tabla 6-7, en donde se señalan las reglas que incumplen las restricciones. Por ejemplo, la regla de asociación 39: $\{Apples\} \Rightarrow \{Soap\}$ incumple RE_2 y, por lo tanto; se descarta, debido a la categoría *Apples* y sus descendientes no pueden estar cercanos a la categoría *Soap and detergents, perfume and toilet preparations* y sus descendientes.

Tabla 6-7: Reglas de asociación resultantes de la Tabla 6-6 que cumplen con las restricciones $RE = \{RE_1, RE_2, RE_3, RE_4, RE_5, RE_6, RE_7\}$.
El conjunto de reglas que cumplen con las restricciones se denomina S'''.

# Rule	Rule	Attribute	W-gain	|R|	UW-score	Support	Confidence	Lift
5	{Apples} ⇒ {Processed liquid milk}	Apples	46	2	35	0.25	1	1
		Processed liquid milk	24					
31	{Apples} ⇒ {Biscuits}	Apples	46	2	31.5	0.167	0.667	0.941
		Biscuits	17					
7	{Apples,Biscuits} ⇒ {Processed liquid milk}	Apples	46	3	29	0.167	1	1
		Biscuits	17					
		Processed liquid milk	24					
32	{Apples,Processed liquid milk} ⇒ {Biscuits}	Apples	46	3	29	0.167	0.667	0.941
		Processed liquid milk	24					
		Biscuits	17					
1	{Biscuits} ⇒ {Processed liquid milk}	Biscuits	17	2	20.5	0.708	1	1
		Processed liquid milk	24					
30	{Processed liquid milk} ⇒ {Biscuits}	Processed liquid milk	24	2	20.5	0.708	0.708	1
		Biscuits	17					
4	{Brooms - brushes} ⇒	Brooms - brushes	10	2	17	0.25	1	1

	{Processed liquid milk}	*Processed liquid milk*	24					
6	*{Bottled waters} ⇒ {Processed liquid milk}*	*Bottled waters*	6	2	15	0.167	1	1
		Processed liquid milk	24					
10	*{Oranges} ⇒ {Processed liquid milk}*	*Oranges*	6	2	15	0.125	1	1
		Processed liquid milk	24					

6.5 Paso 4: Generar grupos de productos cercanos

A partir del conjunto de reglas de asociación encontradas en el paso anterior, i.e., S''', se procede a agrupar las reglas con el fin de reducir la cantidad de reglas obtenidas y de obtener los g grupos de los productos que deben de ir cercanos.

El *clustering* es el proceso de examinar una colección de "puntos" y agrupar los puntos en grupos según alguna medida de distancia. El objetivo es que los puntos de un mismo *cluster* tengan una distancia pequeña entre ellos (según un umbral dado), mientras que los puntos de los distintos *clusters* se encuentran a gran distancia [45].

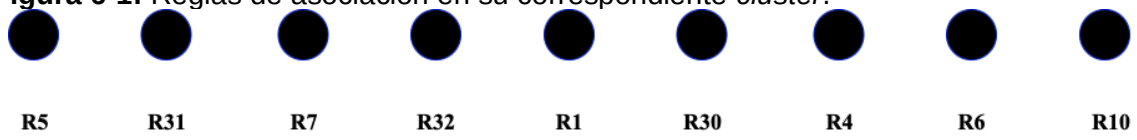
Para obtener los grupos se aplica el algoritmo de agrupación jerárquica (*hierarchical clustering)*. El algoritmo inicia con cada punto formando su propio *cluster* y se resume en los siguientes pasos [45]:

WHILE it is not time to stop DO
    pick the best two clusters to merge;
    combine those two clusters into one cluster;
END;

Como regla de parada se escoge que el algoritmo termine cuando haya g grupos.

Usando como ejemplo el conjunto de reglas $S''' = \{R_5''', R_{31}''', R_7''', R_{32}''', R_1''', R_{30}''', R_4''', R_6''', R_{10}'''\}$ obtenidos en la Tabla 6-7 y con $g = 4$ (i.e., 4 secciones de la tienda), se inicia asignando cada regla a un *cluster*. Lo cual se representa en la Figura 6-1.

Figura 6-1: Reglas de asociación en su correspondiente *cluster*.



La regla de parada indica que se debe llegar a 4 grupos, actualmente se tienen 9, entonces el algoritmo continúa y se deben escoger los dos mejores *clusters* para ser combinados, para esto se calculan las distancias entre cada par de ellos y se selecciona el par con menor distancia.

Generalmente, en el algoritmo de agrupación jerárquica se suele asumir que el espacio es euclidiano, lo que permite representar un *cluster* por un centroide [45], lo cual no aplica aquí, ya que se necesita calcular la distancia entre reglas de asociación que son representadas en un espacio no euclidiano. Para esto se debe usar alguna medida de distancia que se calcule a partir de las reglas.

Debido que no se pueden combinar los puntos de un *cluster* cuando el espacio no es euclidiano, una opción es elegir uno de los puntos del propio *cluster* para representarlo. Idealmente, este punto debería estar cerca de todos los puntos del *cluster*, por lo que en cierto sentido se encuentra en el "centro". Se denomina a este punto representativo el *clustroid*. Se puede seleccionar el *clustroid* de varias maneras, cada una de ellas diseñada para minimizar las distancias entre el *clustroid* y los otros puntos del *cluster*. Las opciones más comunes incluyen el punto que minimiza la suma de las distancias a los demás puntos del *cluster*, la distancia máxima a otro punto del *cluster* y la suma de los cuadrados de las distancias a los otros puntos del *cluster* [45]

Para representar las distancias entre reglas se usa la matriz de distancias P, la cual contiene las distancias para todos los pares de las k reglas. También se conoce como matriz de proximidad. P suele es una matriz simétrica y sus elementos diagonales son iguales al valor mínimo de la distancia d [46]. En esta matriz cada elemento (i, j) es igual a la distancia entre reglas R_i, R_j denotada $dist_R(R_i, R_j)$.

La distancia entre dos reglas se calcula así:

A partir de las reglas $R_1: A_1 \Rightarrow B_1$ y $R_2: A_2 \Rightarrow B_2$

$$dist_R(R_1, R_2) = \omega_0 * J(A_1 \cup B_1, A_2 \cup B_2) + \omega_1 * \alpha + \omega_2 * \beta$$

Donde:

$$J(A, B) = \frac{|X \cup Y| - |X \cap Y|}{|X \cup Y|}$$

$$\alpha = 1 - \frac{Support_{R_1} + Support_{R_2}}{2}$$

$$\beta = 1 - \frac{Confidence_{R_1} + Confidence_{R_2}}{2}$$

La distancia de una regla consigo misma, es decir $dist_R(R_i, R_i) = 0$.

En la fórmula anterior se usa la distancia de *Jaccard* $J(X, Y)$ que mide la diferencia entre dos conjuntos por la proporción de elementos diferentes entre los dos conjuntos con respecto a todos los elementos de los dos [40]. Con el fin de dar importancia al soporte y a la confianza en el cálculo de las distancias se definen las variables α y β, en los cuales se promedian los soportes y las confianzas correspondientemente, y se le resta 1 para asignar una menor distancia en los casos con mejor soporte o confianza. Se usan ω_0, ω_1y ω_2 como pesos ponderados, estos permiten dar más importancia a determinados elementos de la fórmula.

Como ejemplo, se calcula la $dist_R(R_5, R_{31})$ con pesos $\omega_0 = 0.6$, $\omega_1 = 0.2$ y $\omega_2 = 0.2$.
Se tiene que $R_5 : \{Apples\} \Rightarrow \{Processed\ liquid\ milk\}$ y $R_{31} : \{Apples\} \Rightarrow \{Biscuits\}$.

$$dist_R(R_5, R_{31}) = 0.6 * J\begin{pmatrix} \{Apples\} \cup \{Processed\ liquid\ milk\}, \\ \{Apples\} \cup \{Biscuits\} \end{pmatrix} + 0.2 * \alpha + \omega_2 * \beta$$

Se calcula la distancia Jaccard:

$$J(\{Apples\} \cup \{Processed\ liquid\ milk\}, \{Apples\} \cup \{Biscuits\})$$

$$= J(\{Apples, Processed\ liquid\ milk\}, \{Apples, Biscuits\})$$

$$= \frac{\left|\begin{matrix}\{Apples, Processed\ liquid\ milk\} \cup \\ \{Apples, Biscuits\}\end{matrix}\right| - \left|\begin{matrix}\{Apples, Processed\ liquid\ milk\} \cap \\ \{Apples, Biscuits\}\end{matrix}\right|}{|\{Apples, Processed\ liquid\ milk\} \cup \{Apples, Biscuits\}|}$$

$$= \frac{|\{Apples, Processed\ liquid\ milk, Biscuits\}| - |\{Apples\}|}{|\{Apples, Processed\ liquid\ milk, Biscuits\}|}$$

$$= \frac{3-1}{3} = \frac{2}{3}$$

Se calcula α y β:

$$\alpha = 1 - \frac{Support_{R_5} + Support_{R_{31}}}{2} = 1 - \frac{0.25 + 0.167}{2} = 0.792$$

$$\beta = 1 - \frac{Confidence_{R_5} + Confidence_{R_{31}}}{2} = 1 - \frac{1 + 0.667}{2} = 0.165$$

Finalmente:

$$dist_R(R_5, R_{31}) = 0.6 * \left(\frac{2}{3}\right) + 0.2 * 0.792 + 0.2 * 0.165 = 0.591$$

Similarmente usando los mismos valores de ω_0, ω_1y ω_2 , se calculan las distancias para todos los pares entre el conjunto de S''', obteniendo la matriz P de la Tabla 6-8, en la cual se omite el cálculo de la matriz triangular inferior ya que $dist_R(R_i, R_j) = dist_R(R_j, R_i)$.

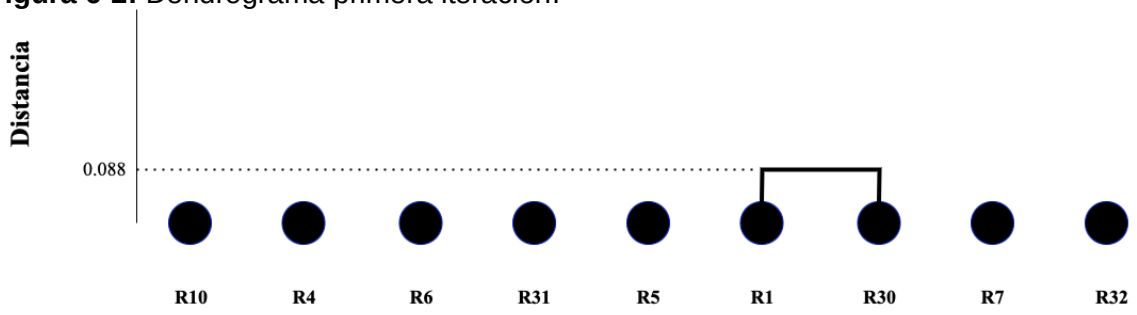
Tabla 6-8: Matriz P de distancias calculadas entre los pares de *cluster*.

	R5	**R31**	**R7**	**R32**	**R1**	**R30**	**R4**	**R6**	**R10**
R5	0	0.592	0.358	0.392	0.504	0.533	0.55	0.558	0.563
R31		0	0.4	0.433	0.546	0.575	0.792	0.8	0.804
R7			0	0.2	0.313	0.342	0.608	0.617	0.621
R32				0	0.346	0.375	0.642	0.65	0.654
R1					0	0.088	0.504	0.513	0.517
R30						0	0.533	0.542	0.546
R4							0	0.558	0.563
R6								0	0.571
R10									0

A partir de esta matriz de distancias P se encuentra que la distancia menor se presenta entre las reglas R_1 y R_{30} que corresponde a 0.088. Por lo cual con las reglas R_1 y R_{30} se unen en un mismo *cluster*, terminando así la primera iteración del algoritmo.

La representación tradicional de la jerarquía de los *clusters* es un dendrograma el cual corresponde a un árbol binario con ejemplos individuales en las hojas (aquí reglas) y un único *cluster* que incluye todos los ejemplos en la raíz. La altura del nodo corresponde al valor del criterio de agrupación en este nodo. Los algoritmos aglomerativos como el algoritmo de agrupación jerárquica que se usó, comienzan en la parte inferior del árbol [47]. En la Figura 6-2 se presenta el dendrograma de la primera iteración**.**

Figura 6-2: Dendrograma primera iteración.



Como el número actual de *clusters* es 8 y la regla de parada es cuando haya 4, se deben seleccionar nuevamente los mejores *clusters* para combinar, para esto es necesario obtener la nueva matriz de distancias considerado el nuevo *cluster*, de la matriz anterior se quitan R_1 y R_{30} y se remplazan por una entrada llamada R_1,R_{30}. En la Tabla 6-9 se presenta la matriz de distancia con las distancias faltantes del nuevo *cluster*.

Tabla 6-9: Matriz de distancias con datos faltantes para el nuevo *cluster* R_1,R_{30}.

.

	R5	R31	R7	R32	R1, R30	R4	R6	R10
R5	0	0.592	0.358	0.392	?	0.55	0.558	0.563
R31		0	0.4	0.433	?	0.792	0.8	0.804
R7			0	0.2	?	0.608	0.617	0.621
R32				0	?	0.642	0.65	0.654
R1, R30					0	?	?	?
R4						0	0.558	0.563
R6							0	0.571
R10								0

En el cálculo de las nuevas distancias con respecto al nuevo *cluster* se usa la función *linkage* que permite medir la distancia entre los *clusters* [48]. Existen diferentes fórmulas de *linkage* para calcular la distancia entre los *cluster* como las presentadas en la Tabla 6-10.

Tabla 6-10: Algunas fórmulas de *linkage* para calcular la distancia entre *cluster*. Adaptada de [49] [50].

Método	**Fórmula**
Single linkage	$d_{c1\rightarrow c2} =$ $min\ \{\ d(i,j) : Donde\ \ i\ está\ en\ el\ cluster\ c1\ y\ el\ objecto\ j\ es\ el\ cluster\ c2\ \}$
Complete linkage	$d_{c1\rightarrow c2} =$ $max\ \{\ d(i,j) : Donde\ \ i\ está\ en\ el\ cluster\ c1\ y\ el\ objecto\ j\ es\ el\ cluster\ c2\ \}$
Average linkage	$d_{c1\rightarrow c2} =$ $average\ \{\ d(i,j) : Donde\ i\ está\ en\ el\ cluster\ c1\ y\ el\ objeto\ j\ el\ el\ cluster\ c2\ \}$

Donde la distancia entre dos *clustroid* c_1 y c_2 se denomina $d_{c1\rightarrow c2}$.

En *hierarchical clustering complete linkage*, se juntan en cada paso los dos *clusters* con la mayor distancia máxima entre pares. El algoritmo de *complete linkage* suele producir jerarquías más útiles que el algoritmo de *single linkage* [47] [51].

En *hierarchical clustering single linkage*, se juntan en cada paso los dos *clusters* cuyos dos miembros tienen la menor distancia entre pares [51].

En *hierarchical clustering* a*verage linkage*, se juntan en cada paso los dos *clusters* cuyos dos miembros tienen la menor distancia promedio entre pares [51].

En este trabajo para ejemplificar se utiliza *hierarchical clustering complete linkage* para calcular las distancias entre los *clusters* obtenidos en la Tabla 6-9. En el caso de la distancia entre el *cluster* $\{R5\}$ y el *cluster* $\{R1, R30\}$ se calcula de la siguiente forma:

$$d_{R5\rightarrow\{R1,R30\}} = max(d_{R5\rightarrow R1}, d_{R5\rightarrow R30}) = (0.504, 0.533) \ = \ 0.533$$

Similarmente se calculan para los demás *cluster* :

$$d_{R31\rightarrow\{R1,R30\}} = max(d_{R31\rightarrow R1}, d_{R5\rightarrow R30}) = (0.546, 0.575) \ = 0.575$$

$$d_{R7\rightarrow\{R1,R30\}} = max(d_{R7\rightarrow R1}, d_{R5\rightarrow R30}) = (0.313, 0.342) \ = 0.342$$

$$d_{R32\rightarrow\{R1,R30\}} = max(d_{R32\rightarrow R1}, d_{R5\rightarrow R30}) = (0.346, 0.375) \ = \ 0.375$$

$$d_{R4\rightarrow\{R1,R30\}} = max(d_{R34\rightarrow R1}, d_{R5\rightarrow R30}) = (0.504, 0.533) \ = \ 0.533$$

$$d_{R6\rightarrow\{R1,R30\}} = max(d_{R6\rightarrow R1}, d_{R5\rightarrow R30}) = (0.513, 0.542) \ = 0.542$$

$$d_{R10\rightarrow\{R1,R30\}} = max(d_{R10\rightarrow R1}, d_{R5\rightarrow R30}) = (0.517, 0.546) \ = 0.546$$

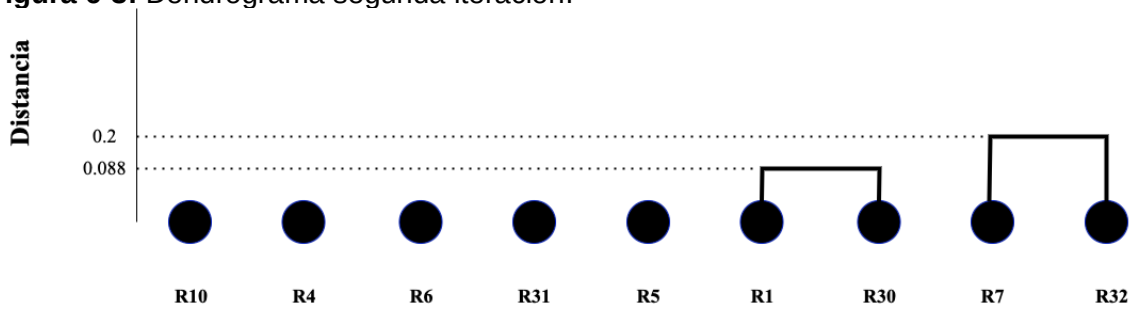
Obteniendo así los valores faltantes de la matriz de distancias presentados en la Tabla 6-11.

Tabla 6-11: Matriz de distancias para el nuevo *cluster* R_1,R_{30}.

	R5	**R31**	**R7**	**R32**	**R1, R30**	**R4**	**R6**	**R10**
R5	0	0.592	0.358	0.392	0.533	0.55	0.558	0.563
R31		0	0.4	0.433	0.575	0.792	0.8	0.804
R7			0	0.2	0.342	0.608	0.617	0.621
R32				0	0.375	0.642	0.65	0.654
R1, R30					0	0.533	0.542	0.546
R4						0	0.558	0.563
R6							0	0.571
R10								0

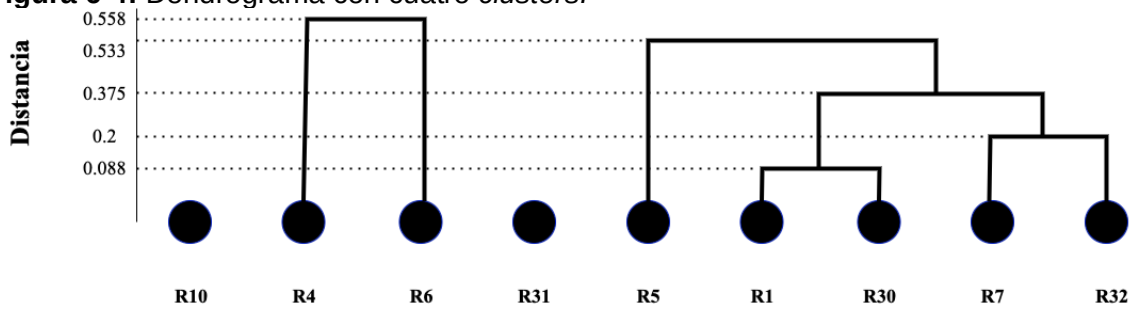
Con esta matriz se encuentra que las reglas R_7 y R_{32} poseen la menor distancia, con $dist_R(R_7, R_{32}) = 0.2$ y se juntan en un mismo *cluster*. En la Figura 6-3 se presenta el dendrograma resultante de la segunda iteración.

Figura 6-3: Dendrograma segunda iteración.



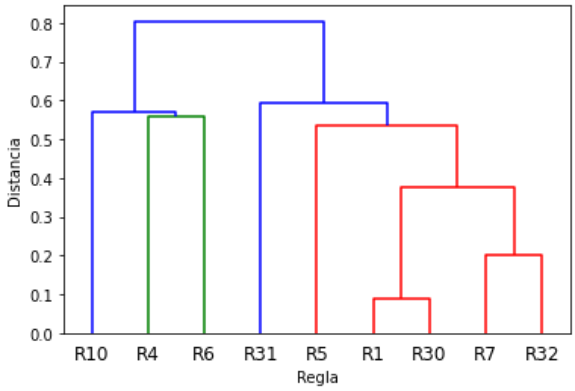
Se continua con los demás pasos y siguientes iteraciones del algoritmo hasta llegar al número indicado de 4 *clusters* donde se cumple la condición de parada y se obtienen los conjuntos de reglas agrupadas. En la Figura 6-4 se muestran los 4 *clusters* de las reglas obtenidos al finalizar el algoritmo.

Figura 6-4: Dendrograma con cuatro *clusters.*



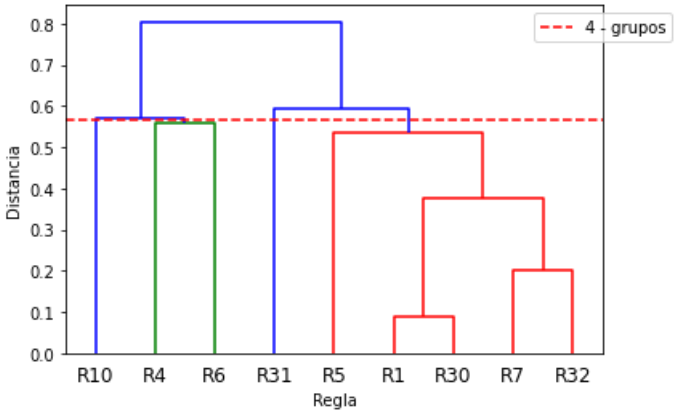
Si se hubiese continuado con el algoritmo y generado el dendrograma completo se obtendría la Figura 6-5. Al trazar una línea horizontal se obtiene el número de *clusters* el cual es igual al número de veces que corte las ramas [47].

Figura 6-5: Dendrograma completo.



Como ejemplo, si se traza en la figura anterior una línea horizontal a una altura de 0.565 se obtienen 4 grupos que corresponden a los mismos de la Figura 6-4, el cual se presenta en la Figura 6-6**.**

Figura 6-6: Dendrograma con cuatro *clusters* calculado a partir del dendrograma completo de la Figura 6-5.



6.6 Paso 5: Generar diseños

Finalmente, en este paso a partir del dendrograma generado en la sección 6.5 y la cantidad g de grupos correspondiente a las secciones de productos de la tienda, se genera el diseño para la tienda. En el ejemplo de la sección 6.5 se obtuvieron los grupos:

- $C_1 = \{R10\}$
- $C_2 = \{R4, R6\}$
- $C_3 = \{R31\}$

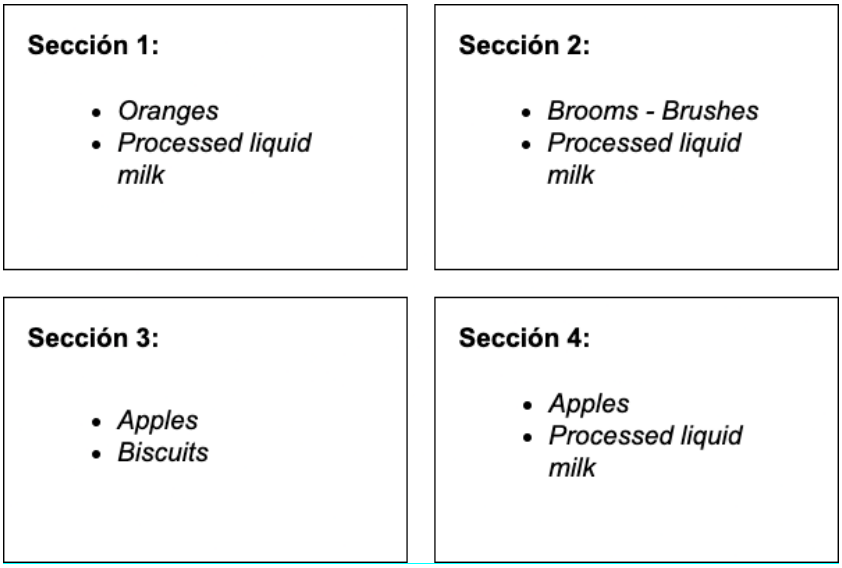
- $C_4 = \{R5,\ R1,\ R30,\ R7,\ R32\}$

Al representar los antecedentes y consecuentes de las reglas obtenidas en un conjunto con los elementos únicos se obtiene:

- $C_1 = \{Oranges, Processed\ liquid\ milk\}$
- $C_2 = \{Brooms - brushes,\ Processed\ liquidmilk,\ Bottled\ watters\}$
- $C_3 = \{Apples, Biscuits\}$
- $C_4 = \{Apples, Processed\ liquid\ milk, Biscuits\}$

En la Figura 6-7 se representan las secciones de la tienda esquemáticamente con los grupos de categorías de productos correspondientes a cada sección.

Figura 6-7: Propuesta de diseño de tienda.

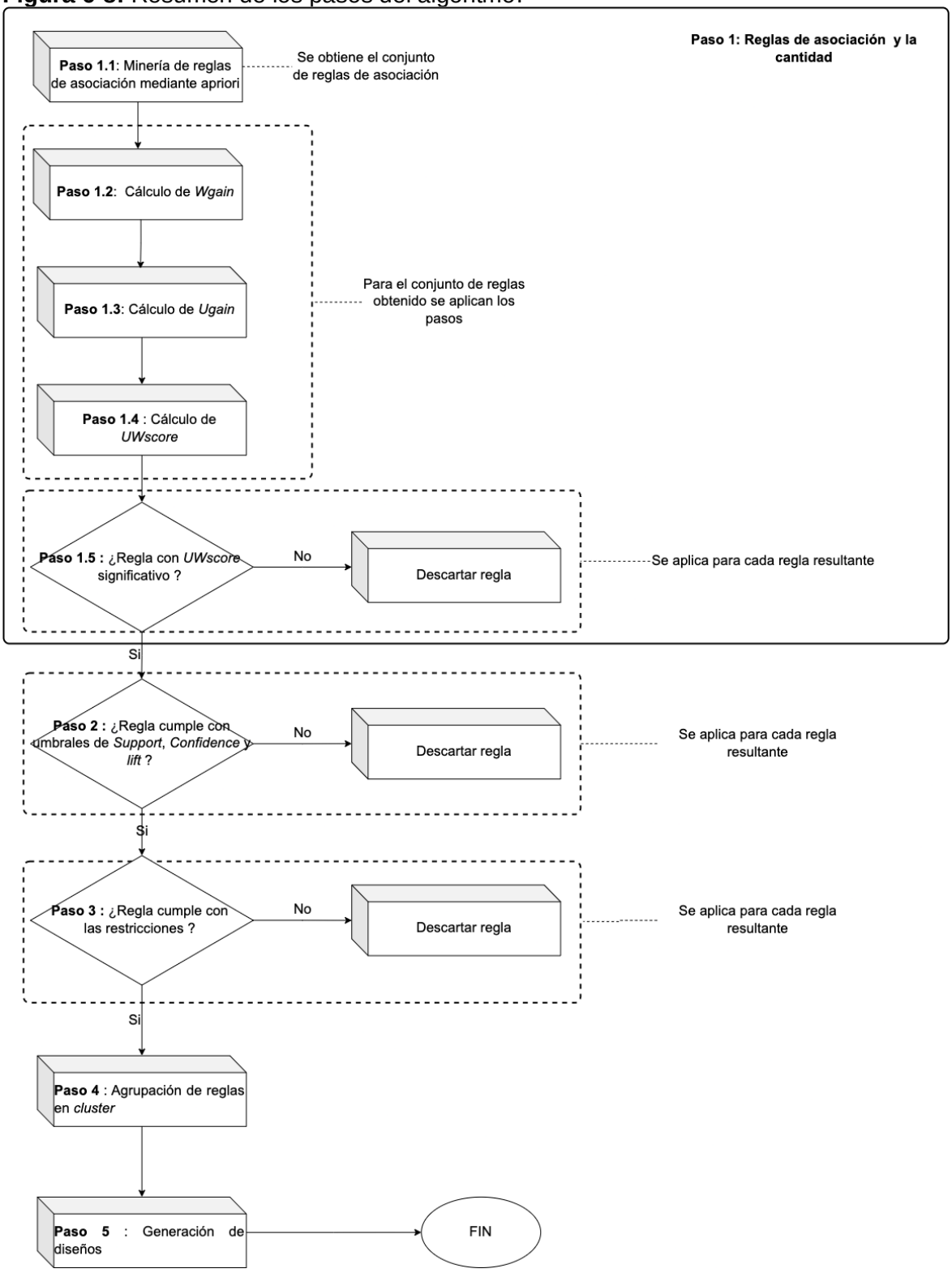


En el diseño anterior se puede notar que existen categorías de productos que se encuentran en varias secciones de la tienda, e.g., *processed liquid milk* se encuentra en las secciones 1, 2 y 4.

6.7 Resumen

Finalmente en la Figura 6-8 se resumen los pasos del algoritmo.

Figura 6-8: Resumen de los pasos del algoritmo.



Si inicialmente no hay datos de las compras de los clientes (*start cold*), e.g., cuando se está iniciando con un supermercado, se puede usar el enfoque tradicional descrito en [1] [24]. Luego, a medida que se generen compras se puede aplicar el algoritmo, aplicar el diseño generado por este y periódicamente aplicar el algoritmo de acuerdo con las compras que se van generando, pudiéndose obtener cambios en las distribuciones de las secciones (evolución de los diseños).

Adicionalmente, el gerente o el analista encargado del diseño podría determinar que algunos productos no van a entrar en el algoritmo porqué tienen una ubicación específica, fija, e.g., cerca de las cajas registradoras se suelen poner libros, revistas o dulces.

7.Experimento

En esta sección se aplican los pasos del algoritmo descritos en la sección 6. Se tomó una muestra de 9835 transacciones (compras) de un supermercado [52], de las cuales se seleccionaron las 200 para ilustrar el algoritmo. En la Tabla 7-1 se muestran 10 transacciones. En el Anexo A se presenta el listado completo.

Tabla 7-1: Muestra de 10 transacciones.

#	***Transaction***
1	*citrus fruit,semi-finished bread,margarine,ready soups*
2	*tropical fruit,yogurt,coffee*
3	*whole milk*
4	*pip fruit,yogurt,cream cheese,meat spreads*
5	*other vegetables,whole milk,condensed milk,long life bakery product*
6	*whole milk,butter,yogurt,rice,abrasive cleaner*
7	*rolls/buns*
8	*other vegetables,UHT-milk,rolls/buns,bottled beer,liquor (appetizer)*
9	*potted plants*
10	*whole milk,cereals*

Nótese que la muestra de datos no incluye la cantidad comprada de cada producto.

Estas transacciones se clasificaron según el árbol de clasificación de productos [35] de la sección 5.1. Las 200 transacciones tienen 118 productos diferentes los cuales fueron manualmente clasificados en *section, division, group, class* y *subclass*. En la Tabla 7-2 se presenta el resultado de la clasificación para 5 productos y en el Anexo B se presenta la clasificación aplicada a los 118 productos.

Tabla 7-2: Muestra de la clasificación de los productos según la propuesta de [35]

.Name	*Section*	*Division*	*Group*	*Class*	*Subclass*
citrus fruit	*Agriculture, forestry and fishery products*	*Products of agriculture, horticulture and market gardening*	*Fruits and nuts*	*Citrus fruits*	*Oranges*
tropical fruit	*Agriculture, forestry and fishery products*	*Products of agriculture, horticulture and market gardening*	*Fruits and nuts*	*Tropical and subtropical fruits*	*Bananas*
whole milk	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Dairy products and egg products*	*Processed liquid milk, cream and whey*	*Processed liquid milk*	*Processed liquid milk*
pip fruit	*Agriculture, forestry and fishery products*	*Products of agriculture, horticulture and market gardening*	*Fruits and nuts*	*Pome fruits and stone fruits*	*Apples*

Se selecciona utilizar la clasificación a nivel de *Group* en las transacciones como entrada al algoritmo. A continuación se describen los resultados obtenidos en cada uno de los pasos.

7.1 Paso 1: Reglas de asociación y la cantidad

7.1.1 Paso 1.1: Minería de reglas de asociación mediante el algoritmo Apriori

A partir de las transacciones del Anexo A se generan las reglas de asociación mediante el algoritmo Apriori con un soporte mínimo de 0.08 y una confianza mínima de 0.3. Al ordenar descendentemente por nivel de confianza se muestran las reglas en la Tabla 7-3.

Tabla 7-3: Reglas de asociación ordenadas descendentemente por *confidence* generadas a partir de las transacciones del Anexo A.

# *Rule*	*Antecedent*	*Consequent*	*Support*	*Confidence*	*lift*
1	*{Fruits and nuts}*	*{Other dairy products}*	0.130	0.542	1.617
2	*{Meat and meat products}*	*{Bakery products}*	0.140	0.528	1.428

3	*{Processed liquid milk, cream and whey}*	*{Other dairy products}*	0.160	0.508	1.516
4	*{Vegetables}*	*{Bakery products}*	0.120	0.480	1.297
5	*{Other dairy products}*	*{Processed liquid milk, cream and whey}*	0.160	0.478	1.516
6	*{Processed liquid milk, cream and whey}*	*{Bakery products}*	0.150	0.476	1.287
7	*{Soft drinks; bottled mineral waters}*	*{Bakery products}*	0.120	0.444	1.201
8	*{Fruits and nuts}*	*{Meat and meat products}*	0.105	0.438	1.651
9	*{Meat and meat products}*	*{Processed liquid milk, cream and whey}*	0.110	0.415	1.318
10	*{Bakery products}*	*{Processed liquid milk, cream and whey}*	0.150	0.405	1.287
11	*{Vegetables}*	*{Processed liquid milk, cream and whey}*	0.100	0.400	1.270
12	*{Meat and meat products}*	*{Fruits and nuts}*	0.105	0.396	1.651
13	*{Fruits and nuts}*	*{Processed liquid milk, cream and whey}*	0.095	0.396	1.257
14	*{Other dairy products}*	*{Fruits and nuts}*	0.130	0.388	1.617
15	*{Vegetables}*	*{Other dairy products}*	0.095	0.380	1.134
16	*{Bakery products}*	*{Meat and meat products}*	0.140	0.378	1.428
17	*{Fruits and nuts}*	*{Bakery products}*	0.090	0.375	1.014
18	*{Other dairy products}*	*{Bakery products}*	0.125	0.373	1.008
19	*{Fruits and nuts}*	*{Vegetables}*	0.085	0.354	1.417
20	*{Soft drinks; bottled mineral waters}*	*{Processed liquid milk, cream and whey}*	0.095	0.352	1.117
21	*{Processed liquid milk, cream and whey}*	*{Meat and meat products}*	0.110	0.349	1.318
22	*{Vegetables}*	*{Fruits and nuts}*	0.085	0.340	1.417
23	*{Vegetables}*	*{Meat and meat products}*	0.085	0.340	1.283
24	*{Meat and meat products}*	*{Other dairy products}*	0.090	0.340	1.014
25	*{Bakery products}*	*{Other dairy products}*	0.125	0.338	1.008
26	*{Bakery products}*	*{Soft drinks; bottled mineral waters}*	0.120	0.324	1.201
27	*{Bakery products}*	*{Vegetables}*	0.120	0.324	1.297
28	*{Meat and meat products}*	*{Vegetables}*	0.085	0.321	1.283
29	*{Processed liquid milk, cream and whey}*	*{Vegetables}*	0.100	0.317	1.270
30	*{Soft drinks; bottled mineral waters}*	*{Other dairy products}*	0.085	0.315	0.940
31	*{Processed liquid milk, cream and whey}*	*{Soft drinks; bottled mineral waters}*	0.095	0.302	1.117

32	*{Processed liquid milk, cream and whey}*	*{Fruits and nuts}*	0.095	0.302	1.257

7.1.2 Paso 1.2: Cálculo de la medida Wgain

Para calcular el *wgain* se requiere la cantidad total comprada de cada producto (*group*) en las 200 transacciones. Como las transacciones de la muestra no incluyen esta cantidad, para obtenerla se generaron números enteros pseudoaleatorios en el rango de 1 - 50 usando el código de Python de la Figura 7-1, donde se usó la semilla 53. En la Tabla 7-4 se muestran las cantidades obtenidas para cada *group*.

Figura 7-1: Fragmento de código en Python para la generación de números enteros pseudoaleatorios.

```
import numpy as np
np.random.seed(53)
print(np.random.randint(low=1, high=50, size=7, dtype=int))
```

Tabla 7-4: Cantidad total obtenida para cada *group*.

Group	***Wgain***
Fruits and nuts	26
Meat and meat products	30
Processed liquid milk, cream and whey	38
Vegetables	26
Other dairy products	49
Soft drinks; bottled mineral waters	25
Bakery products	13

Ahora, se calcula el *wgain* de cada atributo de las reglas de la Tabla 7-3. En la Tabla 7-5 se presentan los resultados para las primeras 5 reglas. En el Anexo C se presentan los resultados para el resto (se usó "*ANY*" para los productos que no se pudieron clasificar en un grupo de los propuestos en [35]).

Tabla 7-5: *Wgain* de las primeras 5 reglas de asociación de la Tabla 7-3.

# Rule	***Rule***	***Attribute***	***W-gain***
1	*{Fruits and nuts}* ⇒ *{Other dairy*	*Fruits and nuts*	26

	products}	*Other dairy products*	49
2	*{Meat and meat products} ⇒ {Bakery products}*	*Meat and meat products*	30
		Bakery products	13
3	*{Processed liquid milk, cream and whey} ⇒ {Other dairy products}*	*Processed liquid milk, cream and whey*	38
		Other dairy products	49
4	*{Vegetables} ⇒ {Bakery products}*	*Vegetables*	26
		Bakery products	13
5	*{Other dairy products} ⇒ {Processed liquid milk, cream and whey}*	*Other dairy products*	49
		Processed liquid milk, cream and whey	38

7.1.3 Paso 1.3: Cálculo de la medida *Ugain*

Según lo explicado en la sección 6.2.5 como no se considera la utilidad de los productos, entonces $Ugain_j$ es igual para todos los atributos (*groups*). Por lo tanto, se omite el cálculo de esta variable ya que no afecta el *UWScore*, ver siguiente paso.

7.1.4 Paso 1.4: Cálculo de *UWscore* a partir de *Wgain* y *Ugain*.

A partir de los pasos anteriores se calcula el *UWScore*. Se seleccionó un umbral de $UWScore \geq 20$ y se ordenan las reglas descendentemente por este atributo. En la Tabla 7-6 se presentan las 5 primeras reglas. En el Anexo C se presenta el resultado para todas las reglas.

Tabla 7-6: *UWScore* de las cinco mejores reglas de asociación de la Tabla 7-3.

# Rule	***Rule***	***Attribute***	***W-gain***	***|R|***	***UW-score***
3	*{Processed liquid milk, cream and whey} => {Other dairy products}*	*Processed liquid milk, cream and whey*	38	2	43.5
		Other dairy products	49		
5	*{Other dairy products} => {Processed liquid milk, cream and whey}*	*Other dairy products*	49	2	43.5
		Processed liquid milk, cream and whey	38		
24	*{Meat and meat products} => {Other dairy products}*	*Meat and meat products*	30	2	39.5
		Other dairy products	49		
1	*{Fruits and nuts} => {Other dairy products}*	*Fruits and nuts*	26	2	37.5
		Other dairy products	49		

14	*{Other dairy products} => {Fruits and nuts}*	*Other dairy products*	49	2	37.5
		Fruits and nuts	26		

7.2 Paso 2: Selección de reglas con *support*, confidence y *lift* significativos

Se seleccionan las reglas del paso anterior que cumplen con los siguientes umbrales soporte mínimo $suppmin = 0.1$, confianza mínima $minconf = 0.4$ y el *lift* mínimo $minlift = 1$. Las reglas resultantes se presentan en la Tabla 7-7.

Tabla 7-7: Reglas de asociación resultantes de la Tabla 7-7 que cumplen con los umbrales $minsupp = 0.1$, $minconf = 0.4$ y $minlift = 1$.

#	Rule	Attribute	W-gain	|R|	UW-score	support	confidence	lift
3	*{Processed liquid milk, cream and whey} ⇒ {Other dairy products}*	*Processed liquid milk, cream and whey*	38	2	43.5	0.160	0.508	1.52
		Other dairy products	49					
5	*{Other dairy products} ⇒ {Processed liquid milk, cream and whey}*	*Other dairy products*	49	2	43.5	0.160	0.478	1.52
		Processed liquid milk, cream and whey	38					
1	*{Fruits and nuts} ⇒ {Other dairy products}*	*Fruits and nuts*	26	2	37.5	0.130	0.542	1.62
		Other dairy products	49					
9	*{Meat and meat products} ⇒ {Processed liquid milk, cream and whey}*	*Meat and meat products*	30	2	34	0.110	0.415	1.32
		Processed liquid milk, cream and whey	38					
11	*{Vegetables} ⇒ {Processed liquid milk, cream and whey}*	*Vegetables*	26	2	32	0.100	0.400	1.27
		Processed liquid milk, cream and whey	38					
8	*{Fruits and nuts} ⇒ {Meat and meat products}*	*Fruits and nuts*	26	2	28	0.105	0.438	1.65
		Meat and meat products	30					
6	*{Processed liquid milk, cream and whey} ⇒ {Bakery products}*	*Processed liquid milk, cream and whey*	38	2	25.5	0.150	0.476	1.29
		Bakery products	13					
10	*{Bakery products} ⇒ {Processed liquid milk, cream and whey}*	*Bakery products*	13	2	25.5	0.150	0.405	1.29
		Processed liquid milk, cream and whey	38					
2	*{Meat and meat products} ⇒ {Bakery products}*	*Meat and meat products*	30	2	21.5	0.140	0.528	1.43
		Bakery products	13					

7.3 Paso 3: Selección de reglas que cumplen las restricciones

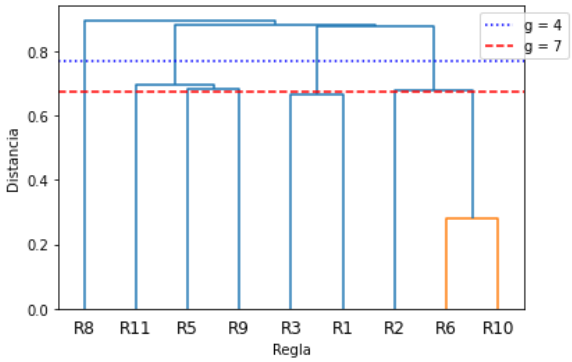
En los resultados del paso anterior (Tabla 7-7) no hay grupos de productos que por concepciones tradicionales o sanitarias no deberían de estar cercanos.

7.4 Paso 4: Generar grupos de productos cercanos

En este paso se consideraron dos escenarios: con $g = 7$ y $g = 4$ grupos de productos, con el fin comparar los diseños generados en cada escenario. En el cálculo de las distancias entre las reglas, se usaron los pesos $\omega_0 = 0.6$, $\omega_1 = 0.2$ y $\omega_2 = 0.2$. Se seleccionaron estos pesos para darle más importancia a una menor distancia *Jaccard*.

Se usó *hierarchical clustering complete linkage* en los cálculos de las distancias de los *clusters*. En la Figura 7-2 se presenta el dendrograma de los grupos obtenidos para $g = 7$ y $g = 4$.

Figura 7-2: Dendrograma de los *clusters* obtenidos.



Para $g = 7$ se obtuvieron los grupos :

- $C_1 = \{R8\}$
- $C_2 = \{R11\}$
- $C_3 = \{R5\}$
- $C_4 = \{R9\}$
- $C_5 = \{R3, R1\}$
- $C_6 = \{R2\}$
- $C_7 = \{R6, R10\}$

Para $g = 4$ se obtuvieron los grupos :

- $C_1 = \{R8\}$
- $C_2 = \{R11,\ R5,\ R9\}$
- $C_3 = \{R3, R1\}$
- $C_4 = \{R2,\ R6,\ R10\}$

7.5 Paso 5: Generar diseños

Finalmente, en este paso a partir de los *clusters* obtenidos en el paso anterior se generan las propuestas de diseño.

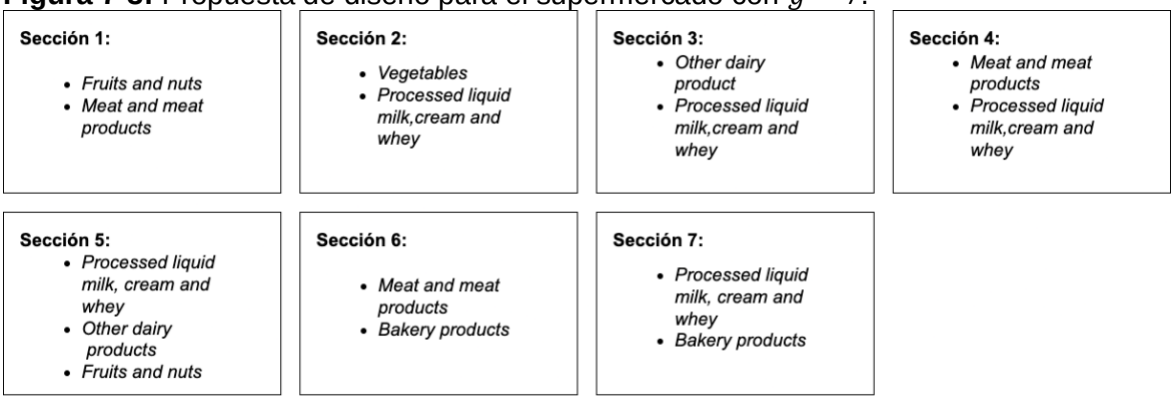
Al representar los antecedentes y consecuentes de las reglas obtenidas en un conjunto se obtiene:

Para $g = 7$:

- $C_1 = \{"Fruits\ and\ nuts", "Meat\ and\ meat\ products"\}$
- $C_2 = \{"Vegetables", "Processed\ liquid\ milk, cream\ and\ whey"\}$
- $C_3 = \{"Other\ dairy\ product", "Processed\ liquid\ milk, cream\ and\ whey"\}$
- $C_4 = \{"Meat\ and\ meat\ products", "Processed\ liquid\ milk, cream\ and\ whey"\}$
- $C_5 = \{"Processed\ liquid\ milk,\ cream\ and\ whey", "Other\ dairy\ products", "Fruits\ and\ nuts"\}$
- $C_6 = \{"Meat\ and\ meat\ products", "Bakery\ products"\}$
- $C_7 = \{"Processed\ liquid\ milk,\ cream\ and\ whey", "Bakery\ products"\}$

La propuesta de diseño para $g = 7$ se muestra en la Figura 7-3.

Figura 7-3: Propuesta de diseño para el supermercado con $g = 7$.

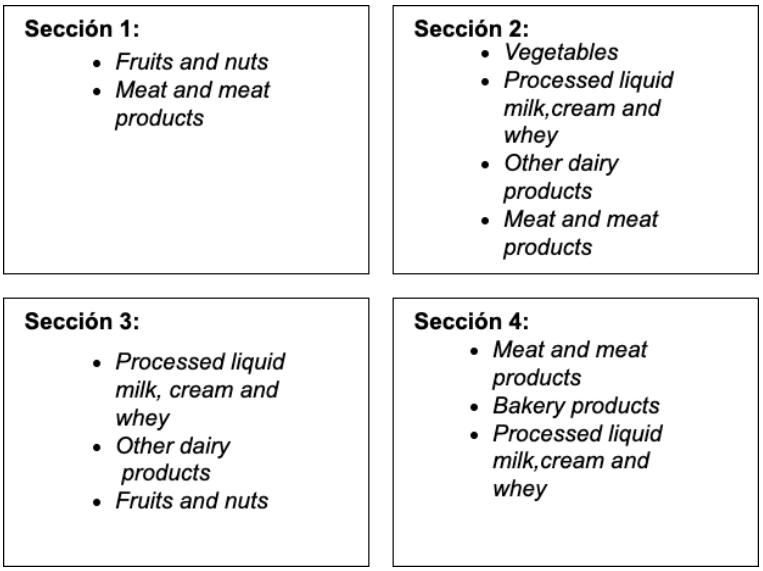


Para $g = 4$:

- $C_1 = \{“Fruits\ and\ nuts”, “Meat\ and\ meat\ products”\}$
- $C_2 = \{“Vegetables”, "Processed\ liquid\ milk,\ cream\ and\ whey",$
  $"Other\ dairy\ products", "Meat\ and\ meat\ products"\}$
- $C_3 = \{"Processed\ liquid\ milk, cream\ and\ whey",$
  $"Other\ dairy\ products", "Fruits\ and\ nuts"\}$
- $C_4 = \{"Meat\ and\ meat\ products", "Bakery\ products",$
  $"Processed\ liquid\ milk", "cream\ and\ whey"\}$

La propuesta de diseño para $g = 4$ se muestra en la Figura 7-4.

Figura 7-4: Propuesta de diseño para el supermercado con $g = 4$.



La muestra de los datos se puede considerar pequeña (200 transacciones) y se pueden encontrar resultados no ajustados a la realidad, e.g., se esperaba que por concepciones tradicionales *Fruits and nuts* y *Vegetables* estuviesen en una misma sección. Sin embargo, en ninguno de los dos diseños ocurrió esto. Además nótese que:

- Las secciones 1 en los dos diseños son iguales.
- La sección 5 del diseño con $g = 7$ es igual a la sección 3 del diseño con $g = 4$.
- Las secciones 2, 3 y 4 con $g = 7$ equivalen a la sección 2 del diseño con $g = 4$.
- Las secciones 6 y 7 con $g = 7$ equivalen a la sección 4 del diseño con $g = 4$.
- Existen categorías de productos que están en más de una sección.

Con una muestra de datos mayor se podrían obtener resultados que posiblemente reflejen algunas concepciones tradicionales o que sugieran configuraciones no intuitivas.

8. Conclusiones y resultados

En esta tesis se analizaron los principales enfoques usados en el diseño de supermercados basados en reglas de asociación, mediante un proceso de revisión y de clasificación sistemática de las investigaciones relevantes donde se recopilaron trabajos enfocados en i) las reglas de asociación donde se considera la cantidad de unidades compradas de cada producto, ii) el diseño de instalaciones usando reglas de asociación iii) reglas de asociación que se enfocan en el diseño de supermercados y finalmente iv) enfoques de diseño de supermercados.

Se propuso un algoritmo con un ejemplo práctico y un experimento para la generación de múltiples diseños de tiendas basados en reglas de asociación, en el cual se considera i) la cantidad en la que un producto es adquirido en la generación de las reglas, ii) el uso del esquema de categorización de productos diseñado por el Departamento de Asuntos Económicos y Sociales de las Naciones Unidas [35], iii) un conjunto de restricciones entre las categorías de productos que no deben situarse cerca (en una misma sección) por leyes o tradiciones y iv) la generación de *clusters* de reglas de asociación usando *hierarchical clustering* con la propuesta de una forma para calcular distancia entre reglas de asociación en la que se aplica la distancia de *Jaccard* y una serie de pesos que le dan importancia a las reglas.

En los resultados obtenidos del experimento, se puede observar que se generaron grupos de productos que contienen elementos que están en más de una sección, e.g, *Processed liquid milk,* lo cual es positivo debido a lo sugerido en la sección 5.1 que un producto (o su categorización) puede estar en una o más secciones de la tienda.

La muestra de datos puede ser considerada como pequeña con respecto al conjunto de datos (alrededor de 2% del total de la muestra) y se pudieron haber encontrado

resultados no ajustados a la realidad. Se esperaba encontrar por concepciones tradicionales que *Fruits and nuts* y *Vegetables* estuviesen en una misma sección. Sin embargo, en ninguno de los dos diseños ocurrió esto.

El algoritmo propuesto se puede extender y se podría aplicar en el diseño de otros espacios físicos, e.g., tiendas de ropa, de productos electrónicos, bodegas, entre otros.

Como trabajo futuro, el algoritmo también se puede extender considerando la variable de utilidad de los productos, factor que fue omitido al calcular *Ugain* en el paso 1.4. de la sección 6.2.4. También se pretende desarrollar una metodología en la que se definan métricas, i.e., que permitan evaluar con cuál diseño se obtiene mayores beneficios monetarios y mejor rendimiento computacional, y comparar las propuestas de diseño obtenidas frente a otras opciones, como el diseño tradicional de supermercados basado en departamentos.

Adicionalmente se pretende desarrollar un módulo de visualización de tal forma que el gerente o un analista encargado pueda: i) ingresar el mapa del supermercado con las secciones disponibles, ii) cargar las transacciones para generar los diseños a partir del algoritmo presentado y iii) un módulo que le genere el mapa con los posibles diseños mostrando los cambios entre las actualizaciones de los diseños a raíz de las nuevas transacciones del supermercado.

8.1 Resultados

A continuación se listan los resultados obtenidos en esta tesis:

- El algoritmo para la generación de múltiples diseños de tiendas basados en reglas de asociación y otros aspectos desarrollado en la sección 6.
- Una nueva forma de calcular la distancia entre reglas de asociación presentada en la sección 6.5.
- Ponencia presentada en el XIV Congreso de Computación para el Desarrollo (COMPDES 2021), organizado por la Universidad Nacional de Ingeniería, la Red Centroamericana de Computación para el Desarrollo y la Universidad de Alcalá de España, en modo virtual desde la ciudad de Managua, Nicaragua, los días 28,

29 y 30 de julio y publicado en el libro de actas del XIV Congreso COMPDES 2021. Título: Análisis y oportunidades para el diseño de supermercados basado en reglas de asociación. ISBN: 978-84-18979-54-5 [53].

- Ponencia presentada en el XV Congreso de Computación para el Desarrollo (COMPDES 2022), organizado por la Universidad Nacional Autónoma de Honduras, la Red Centroamericana de Computación para el Desarrollo y la Universidad de Alcalá de España, en modo virtual desde la ciudad de Tegucigalpa, Honduras, los días 27, 28 y 29 de julio y publicado en el libro de actas del XV Congreso COMPDES 2022. Título: Propuesta esquemática para la generación de múltiples diseños de tiendas basados en reglas de asociación. ISBN: pendiente.

A. Anexo: Transacciones experimento.

#	Transaction	#	Transaction
1	*citrus fruit,semi-finished bread,margarine,ready soups*	**101**	*soda,misc. beverages*
2	*tropical fruit,yogurt,coffee*	**102**	*sausage,pork,grapes,whole milk,rolls/buns,pastry,soda,specialty bar,bathroom cleaner*
3	*whole milk*	**103**	*frankfurter,rolls/buns,bottled water*
4	*pip fruit,yogurt,cream cheese,meat spreads*	**104**	*sausage,whole milk,yogurt,coffee,fruit/vegetable juice,bottled beer,softener,napkins,photo/film,shopping bags*
5	*other vegetables,whole milk,condensed milk,long life bakery product*	**105**	*soda*
6	*whole milk,butter,yogurt,rice,abrasive cleaner*	**106**	*rolls/buns,fruit/vegetable juice*
7	*rolls/buns*	**107**	*beef,citrus fruit,berries,root vegetables,brown bread,detergent*
8	*other vegetables,UHT-milk,rolls/buns,bottled beer,liquor (appetizer)*	**108**	*whole milk,whipped/sour cream,rolls/buns,soda,waffles,candy,cling film/bags*
9	*potted plants*	**109**	*processed cheese*
10	*whole milk,cereals*	**110**	*tropical fruit,pip fruit*
11	*tropical fruit,other vegetables,white bread,bottled water,chocolate*	**111**	*frankfurter,sausage,root vegetables,other vegetables,rolls/buns,oil,pickled vegetables,bottled beer,chewing gum,dishes*
12	*citrus fruit,tropical fruit,whole milk,butter,curd,yogurt,flour,bottled water,dishes*	**112**	*packaged fruit/vegetables,bottled water*
13	*beef*	**113**	*whole milk,dessert,frozen vegetables,rolls/buns,margarine,bottled water,red/blush wine,newspapers*
14	*frankfurter,rolls/buns,soda*	**114**	*berries,whipped/sour cream*
15	*chicken,tropical fruit*	**115**	*citrus fruit,berries,UHT-milk,processed cheese*
16	*butter,sugar,fruit/vegetable juice,newspapers*	**116**	*tropical fruit,whole milk,yogurt*
17	*fruit/vegetable juice*	**117**	*hamburger meat,whole milk,frozen vegetables,rolls/buns,bottled beer,sparkling wine*
18	*packaged fruit/vegetables*	**118**	*hamburger meat,other vegetables,whole milk,frozen vegetables,domestic eggs,soda,dishes*
19	*chocolate*	**119**	*whole milk,cream cheese,soda*
20	*specialty bar*	**120**	*fish,other vegetables,pastry,margarine,coffee,long life bakery product,male cosmetics,shopping bags*
21	*other vegetables*	**121**	*frankfurter,sausage,long life bakery product,waffles*
22	*butter milk,pastry*	**122**	*packaged fruit/vegetables,dessert,semi-finished bread,canned beer*
23	*whole milk*	**123**	*pastry,soda*
24	*tropical fruit,cream cheese,processed cheese,detergent,newspapers*	**124**	*other vegetables,red/blush wine*

25	*tropical fruit,root vegetables,other vegetables,frozen dessert,rolls/buns,flour,sweet spreads,salty snack,waffles,candy,bathroom cleaner*	**125**	*misc. beverages*
26	*bottled water,canned beer*	**126**	*frankfurter,sausage,citrus fruit,tropical fruit,pip fruit,whole milk,hard cheese,cream cheese,domestic eggs,rice,sugar,canned fruit,misc. beverages,fruit/vegetable juice,hygiene articles*
27	*yogurt*	**127**	*sparkling wine*
28	*sausage,rolls/buns,soda,chocolate*	**128**	*specialty bar*
29	*other vegetables*	**129**	*sausage,butter milk,rolls/buns*
30	*brown bread,soda,fruit/vegetable juice,canned beer,newspapers,shopping bags*	**130**	*curd cheese*
31	*yogurt,beverages,bottled water,specialty bar*	**131**	*frankfurter,dessert*
32	*hamburger meat,other vegetables,rolls/buns,spices,bottled water,hygiene articles,napkins*	**132**	*whole milk,yogurt,beverages,brown bread,shopping bags*
33	*root vegetables,other vegetables,whole milk,beverages,sugar*	**133**	*newspapers*
34	*pork,berries,other vegetables,whole milk,whipped/sour cream,artif. sweetener,soda,abrasive cleaner*	**134**	*berries,soda*
35	*beef,grapes,detergent*	**135**	*curd,flour,liquor (appetizer)*
36	*pastry,soda*	**136**	*tropical fruit,other vegetables,coffee*
37	*fruit/vegetable juice*	**137**	*soda*
38	*canned beer*	**138**	*rolls/buns,Instant food products*
39	*root vegetables,other vegetables,whole milk,dessert*	**139**	*canned beer*
40	*citrus fruit,zwieback,newspapers*	**140**	*canned beer*
41	*sausage,rolls/buns,soda,canned beer,specialty bar,shopping bags*	**141**	*frankfurter,sausage,whole milk,rolls/buns,margarine*
42	*tropical fruit,root vegetables,whole milk,yogurt,domestic eggs,brown bread,pastry,sugar,cereals,coffee,soda,waffles,candy*	**142**	*other vegetables,misc. beverages,canned beer*
43	*berries,yogurt*	**143**	*whole milk,frozen vegetables*
44	*canned beer*	**144**	*whole milk*
45	*butter milk,yogurt,cream cheese,spread cheese,rolls/buns,bottled water,soda,newspapers*	**145**	*citrus fruit,butter,cream cheese,white bread,oil,margarine,newspapers*
46	*coffee*	**146**	*frankfurter,rolls/buns*
47	*pastry,bottled water*	**147**	*frankfurter,ham,pip fruit,other vegetables,whole milk,soft cheese,frozen vegetables,rolls/buns,sugar,baking powder*
48	*rolls/buns*	**148**	*butter milk,yogurt,honey,soda,detergent*
49	*misc. beverages*	**149**	*hamburger meat,citrus fruit,onions,rolls/buns,bottled water,dental care*
50	*root vegetables,other vegetables,butter,curd,whipped/sour cream,UHT-milk,hard cheese,rolls/buns,bottled water,long life bakery product*	**150**	*pasta,popcorn*
51	*sausage,rolls/buns,cat food,newspapers*	**151**	*other vegetables,whole milk,meat spreads,cake bar,chocolate,shopping bags*
52	*canned beer*	**152**	*sausage,pork,citrus fruit,dessert,white bread,dishes,candles*
53	*ham,grapes,other vegetables,whole milk*	**153**	*rolls/buns*
54	*turkey,tropical fruit,other vegetables,curd,domestic eggs,brown bread,semi-finished bread,margarine,baking powder,fruit/vegetable juice*	**154**	*grapes,other vegetables,curd,dessert,yogurt,napkins*
55	*whole milk,yogurt,processed cheese,pickled vegetables,soda*	**155**	*root vegetables,other vegetables,whole milk,yogurt,beverages,soft cheese,frozen vegetables,fruit/vegetable juice,bottled beer*
56	*whole milk,curd,yogurt,pastry*	**156**	*coffee,shopping bags*

57	*packaged fruit/vegetables,brown bread,canned beer*	**157**	*popcorn*
58	*rolls/buns,oil,bottled water,chewing gum,chocolate marshmallow,hygiene articles,napkins*	**158**	*whole milk,rolls/buns,sugar,bottled water,soda*
59	*ham,beef,whipped/sour cream,ice cream,rolls/buns,cat food*	**159**	*tropical fruit,other vegetables,packaged fruit/vegetables,whole milk,snack products*
60	*rolls/buns,pastry,sugar*	**160**	*beef,herbs,flour,cling film/bags,flower soil/fertilizer*
61	*other vegetables,whole milk,frozen vegetables,canned fish,salty snack,seasonal products,detergent*	**161**	*whole milk,butter,dessert,rolls/buns*
62	*sausage,pastry*	**162**	*frankfurter,soda,chewing gum*
63	*sausage,beef,whole milk*	**163**	*tropical fruit,specialty cheese,bottled water*
64	*frankfurter,tropical fruit,rolls/buns,brown bread,sugar*	**164**	*canned beer*
65	*rolls/buns,pastry,soda*	**165**	*finished products,other vegetables,frozen vegetables,bottled beer*
66	*whole milk*	**166**	*beverages,soda*
67	*curd cheese,coffee*	**167**	*hamburger meat,root vegetables,whole milk*
68	*red/blush wine,newspapers*	**168**	*whole milk,pastry,long life bakery product*
69	*sausage,whole milk,curd*	**169**	*finished products,white wine,shopping bags*
70	*tropical fruit,pip fruit,berries,whole milk,frozen potato products,rolls/buns,pickled vegetables,chocolate*	**170**	*other vegetables*
71	*red/blush wine*	**171**	*ham,tropical fruit,other vegetables,whole milk,white bread,brown bread,fruit/vegetable juice,bottled beer*
72	*whole milk,butter,margarine,specialty fat,specialty chocolate,candles,flower (seeds)*	**172**	*other vegetables*
73	*frankfurter,citrus fruit,whole milk,domestic eggs,oil,sparkling wine,specialty chocolate,newspapers*	**173**	*curd,butter milk,cocoa drinks,salty snack,popcorn*
74	*whole milk,meat spreads,soda*	**174**	*bottled beer*
75	*frozen potato products*	**175**	*curd,dessert,soda,salty snack,waffles,cake bar,chocolate,shopping bags*
76	*other vegetables,whole milk,rolls/buns,sugar*	**176**	*whole milk,cream cheese,cereals,dog food*
77	*citrus fruit,whole milk,curd,butter milk,pastry*	**177**	*whole milk,salt*
78	*flour,salt,bottled water,fruit/vegetable juice,bottled beer*	**178**	*dessert,ice cream*
79	*sugar,bottled water,soda,bottled beer*	**179**	*sausage,soft cheese,coffee*
80	*frozen meals,coffee*	**180**	*other vegetables,bottled water,waffles,dishes,shopping bags*
81	*chocolate*	**181**	*ham,whole milk,rolls/buns*
82	*root vegetables,whole milk,frozen vegetables,domestic eggs,rolls/buns,oil,margarine,canned vegetables,fruit/vegetable juice,bathroom cleaner*	**182**	*other vegetables,ice cream*
83	*root vegetables,onions,hard cheese,frozen vegetables*	**183**	*other vegetables,bottled water*
84	*herbs,condensed milk,frozen vegetables,salt,oil,white wine,candles*	**184**	*soda,bottled beer,white wine,prosecco*
85	*bottled water*	**185**	*herbs,candy*
86	*sausage,root vegetables,onions,yogurt,rolls/buns,brandy,detergent*	**186**	*frankfurter,ham,tropical fruit,pip fruit,root vegetables,other vegetables,whole milk,butter,curd,yogurt,whipped/sour cream,soft cheese,sliced cheese,cream cheese,frozen meals,frozen fish,domestic eggs,white bread,sugar,cereals,fruit/vegetable juice,bottled beer,waffles*
87	*coffee,newspapers*	**187**	*rolls/buns,canned beer*
88	*pork,other vegetables,whipped/sour cream*	**188**	*curd,cream cheese,semi-finished bread,zwieback,cake bar,specialty bar,shopping bags*

89	*berries,yogurt*	**189**	*frankfurter,tropical fruit,whole milk,spread cheese,margarine,shopping bags*
90	*soda*	**190**	*other vegetables*
91	*berries*	**191**	*dog food*
92	*fruit/vegetable juice,salty snack,candles*	**192**	*specialty bar*
93	*citrus fruit,butter milk,yogurt,cream cheese*	**193**	*curd,salt,liquor (appetizer),detergent,shopping bags*
94	*pork,hamburger meat,citrus fruit,berries,onions,whipped/sour cream,rice,coffee,chocolate*	**194**	*frankfurter,rolls/buns*
95	*detergent*	**195**	*soda*
96	*grapes,photo/film*	**196**	*beef,berries,root vegetables,whole milk,dessert,spread cheese,brown bread*
97	*sausage,sliced cheese,rolls/buns,brown bread,fruit/vegetable juice,canned beer,red/blush wine,waffles,shopping bags*	**197**	*canned beer*
98	*chicken,hamburger meat,citrus fruit,tropical fruit,root vegetables,other vegetables,butter,pasta,oil,detergent*	**198**	*pork,beef,pip fruit,herbs,spices*
99	*whole milk,yogurt,brown bread*	**199**	*frankfurter,citrus fruit,UHT-milk,margarine,sugar*
100	*citrus fruit,tropical fruit*	**200**	*sausage,bottled beer,liquor (appetizer)*

B. Anexo: Categorización de los productos.

Name	***Section***	***Division***	***Group***	***Class***	***Subclass***
citrus fruit	*Agriculture, forestry and fishery products*	*Products of agriculture, horticulture and market gardening*	*Fruits and nuts*	*Citrus fruits*	*Oranges*
tropical fruit	*Agriculture, forestry and fishery products*	*Products of agriculture, horticulture and market gardening*	*Fruits and nuts*	*Tropical and subtropical fruits*	*Bananas*
whole milk	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Dairy products and egg products*	*Processed liquid milk, cream and whey*	*Processed liquid milk*	*Processed liquid milk*
pip fruit	*Agriculture, forestry and fishery products*	*Products of agriculture, horticulture and market gardening*	*Fruits and nuts*	*Pome fruits and stone fruits*	*Apples*
other vegetables	*Agriculture, forestry and fishery products*	*Products of agriculture, horticulture and market gardening*	*Vegetables*	*Vegetables, fresh, n.e.c.*	*Vegetables, fresh, n.e.c.*
rolls/buns	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Grain mill products, starches and starch products; other food products*	*Bakery products*	*Gingerbread and the like; sweet biscuits; waffles and wafers*	*Gingerbread and the like; sweet biscuits; waffles and wafers*
potted plants	*Agriculture, forestry and fishery products*	*Forestry and logging products*	*Non-wood forest products*	*Parts of plants, without flowers or flower buds, and grasses, mosses and lichens, suitable for ornamental purposes*	
beef	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Meat, fish, fruits, vegetables, oils and fats*	*Meat and meat products*	*Meat of mammals, fresh or chilled*	*Meat of cattle, fresh or chilled*
frankfurter	*Food products, beverages*	*Meat, fish, fruits, vegetables,*	*Meat and meat products*	*Preserves and*	*Sausages and similar*

	and tobacco; textiles, apparel and leather products	oils and fats		preparations of meat, meat offal or blood	products of meat, offal or blood
chicken	Food products, beverages and tobacco; textiles, apparel and leather products	Meat, fish, fruits, vegetables, oils and fats	Meat and meat products	Meat of poultry, frozen	Meat of chickens, frozen
butter	Food products, beverages and tobacco; textiles, apparel and leather products	Dairy products and egg products	Other dairy products	Butter and other fats and oils derived from milk	Butter and other fats and oils derived from milk of cattle
fruit/vegetable juice	Food products, beverages and tobacco; textiles, apparel and leather products	Meat, fish, fruits, vegetables, oils and fats	Prepared and preserved fruits and nuts	Fruit juices	Other fruit juices, n.e.c.
packaged fruit/vegetables	Agriculture, forestry and fishery products	Products of agriculture, horticulture and market gardening	Vegetables	Fruit-bearing vegetables	Tomatoes
chocolate	Food products, beverages and tobacco; textiles, apparel and leather products	Grain mill products, starches and starch products; other food products	Cocoa, chocolate and sugar confectionery	Chocolate and other food preparations containing cocoa (except sweetened cocoa powder), in bulk forms	Chocolate and other food preparations containing cocoa (except sweetened cocoa powder), in bulk forms
specialty bar	ANY	ANY	ANY	ANY	ANY
butter milk	Food products, beverages and tobacco; textiles, apparel and leather products	Dairy products and egg products	Processed liquid milk, cream and whey	Whey	Whey
bottled water	Food products, beverages and tobacco; textiles, apparel and leather products	Beverages	Soft drinks; bottled mineral waters	Bottled waters, not sweetened or flavoured	Bottled waters, not sweetened or flavoured
yogurt	Food products, beverages and tobacco; textiles, apparel and leather products	Dairy products and egg products	Other dairy products	Yoghurt and other fermented or acidified milk and cream	Yoghurt and other fermented or acidified milk and cream
sausage	Food products, beverages and tobacco; textiles, apparel and leather products	Meat, fish, fruits, vegetables, oils and fats	Meat and meat products	Preserves and preparations of meat, meat offal or blood	Sausages and similar products of meat, offal or blood
brown bread	Food products, beverages and tobacco; textiles, apparel and leather products	Grain mill products, starches and starch products; other food products	Bakery products	Bread and other bakers' wares	Bread and other bakers' wares
hamburger meat	Food products, beverages and tobacco; textiles, apparel and leather products	Meat, fish, fruits, vegetables, oils and fats	Meat and meat products	Meat of mammals, fresh or chilled	Meat of cattle, fresh or chilled
root vegetables	Agriculture, forestry and fishery products	Products of agriculture, horticulture and market gardening	Vegetables	Root, bulb or tuberous vegetables	Other root, bulb and tuberous vegetables, n.e.c.
pork	Food products, beverages and tobacco; textiles, apparel and leather products	Meat, fish, fruits, vegetables, oils and fats	Meat and meat products	Meat of mammals, fresh or chilled	Meat of pigs, fresh or chilled
pastry	Food products, beverages and tobacco; textiles, apparel and leather products	Grain mill products, starches and starch products; other food products	Bakery products	Pastry goods and cakes	Pastry goods and cakes
canned beer	Food products, beverages and tobacco; textiles, apparel and leather products	Beverages	Malt liquors and malt	Beer made from malt	Beer made from malt
berries	Agriculture, forestry and fishery products	Products of agriculture, horticulture and market gardening	Fruits and nuts	Berries and other fruits	Strawberries
coffee	Food products, beverages and tobacco; textiles, apparel and leather products	Grain mill products, starches and starch products; other food products	Food products n.e.c.	Coffee and tea	Coffee, decaffeinated or roasted
misc. beverages	Food products, beverages and tobacco; textiles, apparel and leather products	Beverages	Soft drinks; bottled mineral waters	Other non-alcoholic caloric beverages	Other non-alcoholic caloric beverages
ham	Food products, beverages and tobacco; textiles, apparel and leather products	Meat, fish, fruits, vegetables, oils and fats	Meat and meat products	Preserves and preparations of meat, meat offal or blood	Pig meat, cuts, salted, dried or smoked (bacon and ham)
turkey	Food products, beverages and tobacco; textiles, apparel and leather products	Meat, fish, fruits, vegetables, oils and fats	Meat of poultry, fresh or chilled	Meat of poultry, fresh or chilled	Meat of turkeys, fresh or chilled
curd cheese	Food products, beverages and tobacco; textiles, apparel and leather products	Dairy products and egg products	Other dairy products	Cheese, fresh or processed	Cheese from milk of cattle, fresh or processed
red/blush wine	Food products, beverages and tobacco; textiles, apparel and leather products	Beverages	Wines	Wine of fresh grapes, whether or not flavoured; grape must	Wine of fresh grapes, except sparkling wine; grape must
frozen potato products	Food products, beverages and tobacco; textiles, apparel and leather products	Meat, fish, fruits, vegetables, oils and fats	Prepared and preserved vegetables, pulses and potatoes	Frozen vegetables, pulses and potatoes	Potatoes, frozen
flour	Food products, beverages and tobacco; textiles,	Grain mill products, starches and starch products; other	Grain mill products	Wheat and meslin flour	Wheat and meslin flour

	apparel and leather products	food products			
sugar	Food products, beverages and tobacco; textiles, apparel and leather products	Grain mill products, starches and starch products; other food products	Sugar and molasses	Refined sugar	Refined sugar
frozen meals	Food products, beverages and tobacco; textiles, apparel and leather products	Meat, fish, fruits, vegetables, oils and fats	Prepared and preserved vegetables, pulses and potatoes	Frozen vegetables, pulses and potatoes	Beans, frozen
herbs	Food products, beverages and tobacco; textiles, apparel and leather products	Grain mill products, starches and starch products; other food products	Food products n.e.c.	Spices and aromatics, processed	Other spices and aromatics, processed
soda	Food products, beverages and tobacco; textiles, apparel and leather products	Beverages	Soft drinks; bottled mineral waters	Bottled waters, not sweetened or flavoured	Bottled waters, not sweetened or flavoured
detergent	Other transportable goods, except metal products, machinery and equipment	Other chemical products; man-made fibres	Soap, cleaning preparations, perfumes and toilet preparations	Soap and detergents, perfume and toilet preparations	Detergents and washing preparations
grapes	Agriculture, forestry and fishery products	Products of agriculture, horticulture and market gardening	Fruits and nuts	Grapes	Grapes
processed cheese	Food products, beverages and tobacco; textiles, apparel and leather products	Dairy products and egg products	Other dairy products	Cheese, fresh or processed	Cheese from milk of cattle, fresh or processed
fish	Food products, beverages and tobacco; textiles, apparel and leather products	Meat, fish, fruits, vegetables, oils and fats	Prepared and preserved fish, crustaceans, molluscs and other aquatic invertebrates	Fish, frozen (excluding fish fillets and fish meat)	Freshwater fish, frozen
sparkling wine	Food products, beverages and tobacco; textiles, apparel and leather products	Beverages	Wines	Wine of fresh grapes, whether or not flavoured; grape must	Sparkling wine of fresh grapes
newspapers	Other transportable goods, except metal products, machinery and equipment	Pulp, paper and paper products; printed matter and related articles	Newspapers and periodicals, daily, in print	Newspapers and periodicals, daily, in print	Newspapers and periodicals, daily, in print
curd	Food products, beverages and tobacco; textiles, apparel and leather products	Dairy products and egg products	Other dairy products	Milk and cream in solid forms	Other milk and cream in solid forms, n.e.c.
pasta	Food products, beverages and tobacco; textiles, apparel and leather products	Grain mill products, starches and starch products; other food products	Macaroni, noodles, couscous and similar farinaceous products	Uncooked pasta, not stuffed or otherwise prepared	Uncooked pasta, not stuffed or otherwise prepared
popcorn	Food products, beverages and tobacco; textiles, apparel and leather products	Grain mill products, starches and starch products; other food products	Grain mill products	Other cereal grain products (including corn flakes)	Other cereal grain products (including corn flakes)
finished products	ANY	ANY	ANY	ANY	ANY
beverages	Food products, beverages and tobacco; textiles, apparel and leather products	Beverages	Soft drinks; bottled mineral waters	Other non-alcoholic caloric beverages	Other non-alcoholic caloric beverages
bottled beer	Food products, beverages and tobacco; textiles, apparel and leather products	Beverages	Malt liquors and malt	Beer made from malt	Beer made from malt
dessert	Food products, beverages and tobacco; textiles, apparel and leather products	Dairy products and egg products	Other dairy products	Ice cream and other edible ice	Ice cream and other edible ice
dog food	Food products, beverages and tobacco; textiles, apparel and leather products	Grain mill products, starches and starch products; other food products	Preparations used in animal feeding; lucerne (alfalfa) meal and pellets	Preparations used in animal feeding	Dog or cat food, put up for retail sale
semi-finished bread	Food products, beverages and tobacco; textiles, apparel and leather products	Grain mill products, starches and starch products; other food products	Bakery products	Bread and other bakers' wares	Bread and other bakers' wares
UHT-milk	Food products, beverages and tobacco; textiles, apparel and leather products	Dairy products and egg products	Processed liquid milk, cream and whey	Processed liquid milk	Processed liquid milk
cereals	Food products, beverages and tobacco; textiles, apparel and leather products	Grain mill products, starches and starch products; other food products	Grain mill products	Other cereal grain products (including corn flakes)	Other cereal grain products (including corn flakes)
cream cheese	Food products, beverages and tobacco; textiles, apparel and leather products	Dairy products and egg products	Other dairy products	Cheese, fresh or processed	Cheese from milk of cattle, fresh or processed
zwieback	Food products, beverages and tobacco; textiles, apparel and leather products	Grain mill products, starches and starch products; other food products	Bakery products	Bread and other bakers' wares	Bread and other bakers' wares
oil	Food products, beverages and tobacco; textiles, apparel and leather products	Meat, fish, fruits, vegetables, oils and fats	Vegetable oils	Sunflower-seed and safflower-seed oil	Sunflower-seed and safflower-seed oil, refined
meat spreads	Food products, beverages and tobacco; textiles, apparel and leather products	Meat, fish, fruits, vegetables, oils and fats	Meat and meat products	Meat of mammals, fresh or chilled	Meat of cattle, fresh or chilled
salt	Ores and minerals;	Other minerals	Salt and pure sodium	Salt and pure sodium	Salt and pure sodium

	electricity, gas and water		*chloride; sea water*	*chloride; sea water*	*chloride; sea water*
onions	*Agriculture, forestry and fishery products*	*Products of agriculture, horticulture and market gardening*	*Vegetables*	*Root, bulb or tuberous vegetables*	*Onions*
condensed milk	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Dairy products and egg products*	*Other dairy products*	*Milk and cream, concentrated or containing added sugar or other sweetening matter, other than in solid forms*	*Condensed milk*
salty snack	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Grain mill products, starches and starch products; other food products*	*Bakery products*	*Crispbread; rusks, toasted bread and similar toasted products*	*Crispbread; rusks, toasted bread and similar toasted products*
photo/film	*Other transportable goods, except metal products, machinery and equipment*	*Furniture; other transportable goods n.e.c.*	*Other manufactured articles n.e.c.*	*Photographic plates, film, paper, paperboard and textiles, exposed*	*Photographic plates, film, paper, paperboard and textiles, exposed but not developed*
sliced cheese	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Dairy products and egg products*	*Other dairy products*	*Cheese, fresh or processed*	*Cheese from milk of cattle, fresh or processed*
whipped/sour cream	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Dairy products and egg products*	*Other dairy products*	*Dairy products n.e.c.*	*Dairy products n.e.c.*
Instant food products	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Grain mill products, starches and starch products; other food products*	*Macaroni, noodles, couscous and similar farinaceous products*	*Pasta, cooked, stuffed or otherwise prepared; couscous*	*Prepared dishes containing stuffed pasta; prepared couscous dishes*
frozen vegetables	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Meat, fish, fruits, vegetables, oils and fats*	*Prepared and preserved vegetables, pulses and potatoes*	*Frozen vegetables, pulses and potatoes*	*Other vegetables and pulses, frozen*
shopping bags	*Other transportable goods, except metal products, machinery and equipment*	*Rubber and plastics products*	*Packaging products of plastics*	*Sacks and bags, of plastics*	*Sacks and bags, of plastics*
specialty cheese	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Dairy products and egg products*	*Other dairy products*	*Cheese, fresh or processed*	*Cheese from milk of cattle, fresh or processed*
white wine	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Beverages*	*Wines*	*Wine of fresh grapes, whether or not flavoured; grape must*	*Wine of fresh grapes, except sparkling wine; grape must*
ice cream	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Dairy products and egg products*	*Other dairy products*	*Ice cream and other edible ice*	*Ice cream and other edible ice*
soft cheese	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Dairy products and egg products*	*Other dairy products*	*Cheese, fresh or processed*	*Cheese from milk of cattle, fresh or processed*
candy	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Grain mill products, starches and starch products; other food products*	*Cocoa, chocolate and sugar confectionery*	*Sugar confectionery (including white chocolate), not containing cocoa; vegetables, fruits, nuts, fruit-peel and other parts of plants, preserved by sugar*	*Sugar confectionery (including white chocolate), not containing cocoa; vegetables, fruits, nuts, fruit-peel and other parts of plants, preserved by sugar*
margarine	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Meat, fish, fruits, vegetables, oils and fats*	*Margarine and similar preparations*	*Margarine and similar preparations*	*Margarine and similar preparations*
white bread	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Grain mill products, starches and starch products; other food products*	*Bakery products*	*Bread and other bakers' wares*	*Bread and other bakers' wares*
cat food	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Grain mill products, starches and starch products; other food products*	*Preparations used in animal feeding; lucerne (alfalfa) meal and pellets*	*Preparations used in animal feeding*	*Dog or cat food, put up for retail sale*
hard cheese	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Dairy products and egg products*	*Other dairy products*	*Cheese, fresh or processed*	*Cheese from milk of cattle, fresh or processed*
candles	*Other transportable goods, except metal products, machinery and equipment*	*Furniture; other transportable goods n.e.c.*	*Other manufactured articles n.e.c.*	*Other articles*	*Articles n.e.c. (including candles, tapers, skins of birds with their feathers, artificial flowers, entertainment articles, hand sieves, hand riddles, vacuum flasks, tailors dummies, animated displays used for shop window*

					dressing, and parts n.e.c.)
long life bakery product	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Grain mill products, starches and starch products; other food products*	*Bakery products*	*Bread and other bakers' wares*	*Bread and other bakers' wares*
liquor (appetizer)	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Beverages*	*Ethyl alcohol; spirits, liqueurs and other spirituous beverages*	*Undenatured ethyl alcohol of an alcoholic strength by volume of less than 80% vol; spirits, liqueurs and other spirituous beverages*	*Other spirituous beverages and undenatured ethyl alcohol of an alcoholic strength by volume of less than 80% vol*
honey	*Agriculture, forestry and fishery products*	*Live animals and animal products (excluding meat)*	*Other animal products*	*Natural honey*	*Natural honey*
chewing gum	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Grain mill products, starches and starch products; other food products*	*Cocoa, chocolate and sugar confectionery*	*Sugar confectionery (including white chocolate), not containing cocoa; vegetables, fruits, nuts, fruit-peel and other parts of plants, preserved by sugar*	*Sugar confectionery (including white chocolate), not containing cocoa; vegetables, fruits, nuts, fruit-peel and other parts of plants, preserved by sugar*
cocoa drinks	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Beverages*	*Soft drinks; bottled mineral waters*	*Other non-alcoholic caloric beverages*	*Other non-alcoholic caloric beverages*
waffles	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Grain mill products, starches and starch products; other food products*	*Bakery products*	*Gingerbread and the like; sweet biscuits; waffles and wafers*	*Gingerbread and the like; sweet biscuits; waffles and wafers*
ready soups	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Meat, fish, fruits, vegetables, oils and fats*	*Prepared and preserved vegetables, pulses and potatoes*	*Other prepared and preserved vegetables, pulses and potatoes*	*Prepared dishes and meals based on vegetables, pulses and potatoes*
rice	*Agriculture, forestry and fishery products*	*Products of agriculture, horticulture and market gardening*	*Cereals*	*Rice*	*Rice, seed*
frozen dessert	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Dairy products and egg products*	*Other dairy products*	*Ice cream and other edible ice*	*Ice cream and other edible ice*
spices	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Grain mill products, starches and starch products; other food products*	*Food products n.e.c.*	*Spices and aromatics, processed*	*Other spices and aromatics, processed*
spread cheese	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Dairy products and egg products*	*Other dairy products*	*Cheese, fresh or processed*	*Cheese from milk of cattle, fresh or processed*
pickled vegetables	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Meat, fish, fruits, vegetables, oils and fats*	*Prepared and preserved vegetables, pulses and potatoes*	*Vegetables, pulses and potatoes, preserved by vinegar or acetic acid*	*Vegetables, pulses and potatoes, preserved by vinegar or acetic acid*
canned fish	*Agriculture, forestry and fishery products*	*Fish and other fishing products*	*Fish live, fresh or chilled for human consumption*	*Freshwater fish, live, fresh or chilled*	*Farmed freshwater fish, live, fresh or chilled*
specialty fat	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Meat, fish, fruits, vegetables, oils and fats*	*Margarine and similar preparations*	*Margarine and similar preparations*	*Margarine and similar preparations*
domestic eggs	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Dairy products and egg products*	*Eggs, in shell, preserved or cooked*	*Eggs, in shell, preserved or cooked*	*Eggs, in shell, preserved or cooked*
cake bar	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Grain mill products, starches and starch products; other food products*	*Bakery products*	*Pastry goods and cakes*	*Pastry goods and cakes*
cling film/bags	*Other transportable goods, except metal products, machinery and equipment*	*Rubber and plastics products*	*Packaging products of plastics*	*Sacks and bags, of plastics*	*Sacks and bags, of plastics*
dishes	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Grain mill products, starches and starch products; other food products*	*Food products n.e.c.*	*Other food products*	*Other prepared dishes and meals*
prosecco	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Beverages*	*Wines*	*Wine of fresh grapes, whether or not flavoured; grape must*	*Wine of fresh grapes, except sparkling wine; grape must*
abrasive cleaner	*Other transportable goods, except metal products, machinery and equipment*	*Other chemical products; man-made fibres*	*Soap, cleaning preparations, perfumes and toilet preparations*	*Soap and detergents, perfume and toilet preparations*	*Detergents and washing preparations*
chocolate marshmallow	*Food products, beverages and tobacco; textiles, apparel and leather products*	*Grain mill products, starches and starch products; other food products*	*Cocoa, chocolate and sugar confectionery*	*Chocolate and other food preparations containing cocoa (except sweetened cocoa powder), in bulk forms*	*Chocolate and other food preparations containing cocoa (except sweetened cocoa powder), in bulk forms*
specialty	*Food products, beverages*	*Grain mill products, starches*	*Cocoa, chocolate and*	*Chocolate and other food*	*Chocolate and other*

chocolate	and tobacco; textiles, apparel and leather products	and starch products; other food products	sugar confectionery	preparations containing cocoa (except sweetened cocoa powder), in bulk forms	food preparations containing cocoa (except sweetened cocoa powder), in bulk forms
snack products	Food products, beverages and tobacco; textiles, apparel and leather products	Meat, fish, fruits, vegetables, oils and fats	Prepared and preserved vegetables, pulses and potatoes	Other prepared and preserved vegetables, pulses and potatoes	Potatoes, otherwise prepared or preserved
flower soil/fertilizer	Other transportable goods, except metal products, machinery and equipment	Basic chemicals	Fertilizers and pesticides	Mineral or chemical fertilizers, nitrogenous	Urea
hygiene articles	Other transportable goods, except metal products, machinery and equipment	Other chemical products; man-made fibres	Soap, cleaning preparations, perfumes and toilet preparations	Soap and detergents, perfume and toilet preparations	Detergents and washing preparations
artif. sweetener	Other transportable goods, except metal products, machinery and equipment	Other chemical products; man-made fibres	Pharmaceutical products	Sugars, chemically pure n.e.c.; sugar ethers and sugar esters and their salts n.e.c.	Sugars, chemically pure n.e.c.; sugar ethers and sugar esters and their salts n.e.c.
seasonal products	ANY	ANY	ANY	ANY	ANY
brandy	Food products, beverages and tobacco; textiles, apparel and leather products	Beverages	Wines	Wine of fresh grapes, whether or not flavoured; grape must	Wine of fresh grapes, except sparkling wine; grape must
dental care	Other transportable goods, except metal products, machinery and equipment	Other chemical products; man-made fibres	Pharmaceutical products	Other pharmaceutical products	Other pharmaceutical products
napkins	Other transportable goods, except metal products, machinery and equipment	Pulp, paper and paper products; printed matter and related articles	Pulp, paper and paperboard	Toilet tissue stock, uncoated kraftliner and other uncoated paper and paperboard; miscellaneous other papers	Toilet or facial tissue stock, towel or napkin stock and similar paper, cellulose wadding and webs of cellulose fibres
sweet spreads	Food products, beverages and tobacco; textiles, apparel and leather products	Grain mill products, starches and starch products; other food products	Cocoa, chocolate and sugar confectionery	Sugar confectionery (including white chocolate), not containing cocoa; vegetables, fruits, nuts, fruit-peel and other parts of plants, preserved by sugar	Sugar confectionery (including white chocolate), not containing cocoa; vegetables, fruits, nuts, fruit-peel and other parts of plants, preserved by sugar
flower (seeds)	Agriculture, forestry and fishery products	Products of agriculture, horticulture and market gardening	Forage products; fibre crops; plants used in perfumery, pharmacy, or for insecticidal, fungicidal or similar purposes; beet, forage plant and flower seeds; natural rubber; living plants, cut flowers and flower buds; unmanufactured tobacco; other raw vegetable materials	Living plants; cut flowers and flower buds; flower seeds	Flower seeds
softener	Other transportable goods, except metal products, machinery and equipment	Other chemical products; man-made fibres	Soap, cleaning preparations, perfumes and toilet preparations	Soap and detergents, perfume and toilet preparations	Detergents and washing preparations
male cosmetics	Other transportable goods, except metal products, machinery and equipment	Other chemical products; man-made fibres	Chemical products n.e.c.	Other chemical products	Other chemical products n.e.c.
canned vegetables	Agriculture, forestry and fishery products	Products of agriculture, horticulture and market gardening	Vegetables	Green leguminous vegetables	Other green leguminous vegetables
baking powder	Food products, beverages and tobacco; textiles, apparel and leather products	Grain mill products, starches and starch products; other food products	Food products n.e.c.	Other food products	Yeasts (active or inactive); other single-cell micro-organisms, dead; prepared baking powders
bathroom cleaner	Other transportable goods, except metal products, machinery and equipment	Other chemical products; man-made fibres	Soap, cleaning preparations, perfumes and toilet preparations	Soap and detergents, perfume and toilet preparations	Detergents and washing preparations
canned fruit	Agriculture, forestry and fishery products	Products of agriculture, horticulture and market gardening	Fruits and nuts	Pome fruits and stone fruits	Apricots
frozen fish	Agriculture, forestry and fishery products	Fish and other fishing products	Fish live, fresh or chilled for human consumption	Freshwater fish, live, fresh or chilled	Farmed freshwater fish, live, fresh or chilled

C. Anexo: Resultados de *WGain*, *UWScore* reglas de asociación.

# Rule	Rule	Attribute	W-gain	|R|	UW-score	support	confidence	lift
3	{Processed liquid milk, cream and whey} ⇒ {Other dairy products}	Processed liquid milk, cream and whey	38	2	43.5	0.160	0.508	1.516
		Other dairy products	49					
5	{Other dairy products} ⇒ {Processed liquid milk, cream and whey}	Other dairy products	49	2	43.5	0.160	0.478	1.516
		Processed liquid milk, cream and whey	38					
24	{Meat and meat products} ⇒ {Other dairy products}	Meat and meat products	30	2	39.5	0.090	0.340	1.014
		Other dairy products	49					
1	{Fruits and nuts} ⇒ {Other dairy products}	Fruits and nuts	26	2	37.5	0.130	0.542	1.617
		Other dairy products	49					
14	{Other dairy products} ⇒ {Fruits and nuts}	Other dairy products	49	2	37.5	0.130	0.388	1.617
		Fruits and nuts	26					
15	{Vegetables} ⇒ {Other dairy products}	Vegetables	26	2	37.5	0.095	0.380	1.134
		Other dairy products	49					
30	{Soft drinks; bottled mineral waters} ⇒ {Other dairy products}	Soft drinks; bottled mineral waters	25	2	37.0	0.085	0.315	0.940
		Other dairy products	49					
9	{Meat and meat products} ⇒ {Processed liquid milk, cream and whey}	Meat and meat products	30	2	34.0	0.110	0.415	1.318
		Processed liquid milk, cream and whey	38					
21	{Processed liquid milk, cream and whey} ⇒ {Meat and meat products}	Processed liquid milk, cream and whey	38	2	34.0	0.110	0.349	1.318
		Meat and meat products	30					
11	{Vegetables} ⇒ {Processed liquid milk, cream and whey}	Vegetables	26	2	32.0	0.100	0.400	1.270
		Processed liquid milk, cream and whey	38					
13	{Fruits and nuts} ⇒ {Processed liquid milk, cream and whey}	Fruits and nuts	26	2	32.0	0.095	0.396	1.257
		Processed liquid milk, cream and whey	38					
29	{Processed liquid milk, cream and whey} ⇒ {Vegetables}	Processed liquid milk, cream and whey	38	2	32.0	0.100	0.317	1.270
		Vegetables	26					
32	{Processed liquid milk, cream and whey} ⇒ {Fruits and nuts}	Processed liquid milk, cream and whey	38	2	32.0	0.095	0.302	1.257
		Fruits and nuts	26					
20	{Soft drinks; bottled mineral waters} ⇒ {Processed liquid milk, cream and whey}	Soft drinks; bottled mineral waters	25	2	31.5	0.095	0.352	1.117
		Processed liquid milk, cream and whey	38					
31	{Processed liquid milk, cream and whey} ⇒ {Soft drinks; bottled mineral waters}	Processed liquid milk, cream and whey	38	2	31.5	0.095	0.302	1.117
		Soft drinks; bottled mineral waters	25					
18	{Other dairy products} ⇒ {Bakery products}	Other dairy products	49	2	31.0	0.125	0.373	1.008

		Bakery products	13					
25	*{Bakery products} ⇒ {Other dairy products}*	*Bakery products*	13	2	31.0	0.125	0.338	1.008
		Other dairy products	49					
8	*{Fruits and nuts} ⇒ {Meat and meat products}*	*Fruits and nuts*	26	2	28.0	0.105	0.438	1.651
		Meat and meat products	30					
12	*{Meat and meat products} ⇒ {Fruits and nuts}*	*Meat and meat products*	30	2	28.0	0.105	0.396	1.651
		Fruits and nuts	26					
23	*{Vegetables} ⇒ {Meat and meat products}*	*Vegetables*	26	2	28.0	0.085	0.340	1.283
		Meat and meat products	30					
28	*{Meat and meat products} ⇒ {Vegetables}*	*Meat and meat products*	30	2	28.0	0.085	0.321	1.283
		Vegetables	26					
19	*{Fruits and nuts} ⇒ {Vegetables}*	*Fruits and nuts*	26	2	26.0	0.085	0.354	1.417
		Vegetables	26					
22	*{Vegetables} ⇒ {Fruits and nuts}*	*Vegetables*	26	2	26.0	0.085	0.340	1.417
		Fruits and nuts	26					
6	*{Processed liquid milk, cream and whey} ⇒ {Bakery products}*	*Processed liquid milk, cream and whey*	38	2	25.5	0.150	0.476	1.287
		Bakery products	13					
10	*{Bakery products} ⇒ {Processed liquid milk, cream and whey}*	*Bakery products*	13	2	25.5	0.150	0.405	1.287
		Processed liquid milk, cream and whey	38					
2	*{Meat and meat products} ⇒ {Bakery products}*	*Meat and meat products*	30	2	21.5	0.140	0.528	1.428
		Bakery products	13					
16	*{Bakery products} ⇒ {Meat and meat products}*	*Bakery products*	13	2	21.5	0.140	0.378	1.428
		Meat and meat products	30					
4	*{Vegetables} ⇒ {Bakery products}*	*Vegetables*	26	2	19.5	0.120	0.480	1.297
		Bakery products	13					
17	*{Fruits and nuts} ⇒ {Bakery products}*	*Fruits and nuts*	26	2	19.5	0.090	0.375	1.014
		Bakery products	13					
27	*{Bakery products} ⇒ {Vegetables}*	*Bakery products*	13	2	19.5	0.120	0.324	1.297
		Vegetables	26					
7	*{Soft drinks; bottled mineral waters} ⇒ {Bakery products}*	*Soft drinks; bottled mineral waters*	25	2	19.0	0.120	0.444	1.201
		Bakery products	13					
26	*{Bakery products} ⇒ {Soft drinks; bottled mineral waters}*	*Bakery products*	13	2	19.0	0.120	0.324	1.201
		Soft drinks; bottled mineral waters	25					

Bibliografía

[1] I. Cil, “Consumption universes based supermarket layout through association rule mining and multidimensional scaling,” *Expert Syst Appl*, vol. 39, no. 10, pp. 8611–8625, Aug. 2012, doi: 10.1016/j.eswa.2012.01.192.

[2] A. Borges, Chairholder, and Auchan, “Toward a new supermarket layout : from industrial categories to one stop shopping organization through a data mining approach,” 2004.

[3] Y. F. Wang, Y. L. Chuang, M. H. Hsu, and H. C. Keh, “A personalized recommender system for the cosmetic business,” *Expert Syst Appl*, vol. 26, no. 3, pp. 427–434, Apr. 2004, doi: 10.1016/J.ESWA.2003.10.001.

[4] M. C. Chen, “Ranking discovered rules from data mining with multiple criteria by data envelopment analysis,” *Expert Syst Appl*, vol. 33, no. 4, pp. 1110–1116, Nov. 2007, doi: 10.1016/J.ESWA.2006.08.007.

[5] Alex. Berson, S. Smith, and Kurt. Thearling, “Building data mining applications for CRM,” p. 510, 2000.

[6] R. Agrawal, T. Imieliński, and A. Swami, “Mining Association Rules between Sets of Items in Large Databases,” in *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, 1993, pp. 207–216. doi: 10.1145/170035.170072.

[7] P. D. McNicholas, T. B. Murphy, and M. O’Regan, “Standardising the lift of an association rule,” *Comput Stat Data Anal*, vol. 52, no. 10, pp. 4712–4721, Jun. 2008, doi: 10.1016/J.CSDA.2008.03.013.

[8] B. A. Kitchenham and S. Charters, “Guidelines for performing Systematic Literature Reviews in Software Engineering,” Jun. 2007. [Online]. Available: https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf

[9] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, “Systematic Mapping Studies in Software Engineering,” in *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering*, 2008, pp. 68–77.

[10] A. Adhikari and P. R. Rao, “Association rules induced by item and quantity purchased,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2008, vol. 4947 LNCS, pp. 478–485. doi: 10.1007/978-3-540-78568-2_37.

[11] F. Alfiah *et al.*, “Data Mining Systems to Determine Sales Trends and Quantity Forecast Using Association Rule and CRISP-DM Method,” *International Journal of Engineering and Techniques*, vol. 4, Accessed: Oct. 22, 2020. [Online]. Available: http://www.ijetjournal.org

[12] P. Kumar and & Ananthanarayana, “Discovery of Frequent Itemsets Based on Minimum Quantity and Support,” 2009. Accessed: Oct. 22, 2020. [Online]. Available: https://www.cscjournals.org/library/manuscriptinfo.php?mc=IJCSS-86

[13] J. Han, J. Pei, and Y. Yin, “Mining frequent patterns without candidate generation,” 2000, pp. 1–12. doi: 10.1145/342009.335372.

[14] J. Han, J. Pei, Y. Yin, and R. Mao, “Mining frequent patterns without candidate generation: A frequent-pattern tree approach,” *Data Min Knowl Discov*, vol. 8, no. 1, pp. 53–87, Jan. 2004, doi: 10.1023/B:DAMI.0000005258.31418.83.

[15] S. Ibrahim and J. Revathy, “A Novel Quantity based Weighted Association Rule Mining,” *International Journal of Engineering Inventions*, vol. 4, no. 3, Aug. 2014.

[16] M. S. Khan, M. Muyeba, and F. Coenen, “A weighted utility framework for mining association rules,” in *Proceedings - EMS 2008, European Modelling Symposium, 2nd UKSim European Symposium on Computer Modelling and Simulation*, 2008, pp. 87–92. doi: 10.1109/EMS.2008.73.

[17] P. S. Sandhu, D. S. Dhaliwal, and S. N. Panda, “Mining utility-oriented association rules: An efficient approach based on profit and quantity,” *International Journal of the Physical Sciences*, vol. 6, no. 2, pp. 301–307, 2011, doi: 10.5897/IJPS09.303.

[18] S. Halim, T. Octavia, and C. Alianto, “Designing facility layout of an amusement arcade using market basket analysis,” in *Procedia Computer Science*, Jan. 2019, vol. 161, pp. 623–629. doi: 10.1016/j.procs.2019.11.165.

[19] “Top 10 FEC & Arcade Game Room Layout Tips | laigames.com.” https://laigames.com/top-10-fec-arcade-game-room-layout-tips/ (accessed Nov. 12, 2020).

[20] Y. L. Chen, K. Tang, R. J. Shen, and Y. H. Hu, “Market basket analysis in a multiple store environment,” *Decis Support Syst*, vol. 40, no. 2, pp. 339–354, Aug. 2005, doi: 10.1016/j.dss.2004.04.009.

[21] L. M. Charlet Annie and A. D. Kumar, “Market Basket Analysis for a Supermarket based on Frequent Itemset Mining,” 2012. Accessed: Oct. 23, 2020. [Online]. Available: www.IJCSI.org

[22] L. C. Annie and D. Ashok Kumar, “Frequent Item set mining for Market Basket Data using K-Apriori algorithm,” *International Journal of Computational Intelligence and Informatics*, vol. 1, no. 1, pp. 14–18, 2011.

[23] M. Ali Alan and A. R. Ince, “Use of Association Rule Mining within the Framework of a Customer-Oriented Approach,” *European Scientific Journal, ESJ*, vol. 12, no. 9, p. 81, Mar. 2016, doi: 10.19044/esj.2016.v12n9p81.

[24] G. R. Peterson, “ECONOMICS IN STORE LAYOUT AND DESIGN.” 1970. Accessed: Nov. 12, 2020. [Online]. Available: https://agris.fao.org/agris-search/search.do?recordID=US2012209216

[25] J. S. Larson, E. T. Bradlow, and P. S. Fader, “An exploratory look at supermarket shopping paths,” *International Journal of Research in Marketing*, vol. 22, no. 4, pp. 395–414, Dec. 2005, doi: 10.1016/j.ijresmar.2005.09.005.

[26] J. Cisewski, “Multivariate Analysis, Clustering, and Classification”.

[27] S. Altuntas, “A novel approach based on utility mining for store layout: A case study in a supermarket,” *Industrial Management and Data Systems*, vol. 117, no. 2, pp. 304–319, 2017, doi: 10.1108/IMDS-01-2016-0040.

[28] S. Altuntas and H. Selim, “Facility layout using weighted association rule-based data mining algorithms: Evaluation with simulation,” *Expert Syst Appl*, vol. 39, no. 1, pp. 3–13, Jan. 2012, doi: 10.1016/j.eswa.2011.06.045.

[29] R. Z. Farahani, M. SteadieSeifi, and N. Asgari, “Multiple criteria facility location problems: A survey,” *Applied Mathematical Modelling*, vol. 34, no. 7. Elsevier, pp. 1689–1709, Jul. 01, 2010. doi: 10.1016/j.apm.2009.10.005.

[30] F. Yener and H. R. Yazgan, “Optimal warehouse design: Literature review and case study application,” *Comput Ind Eng*, vol. 129, pp. 1–13, Mar. 2019, doi: 10.1016/j.cie.2019.01.006.

[31] J. C. H. Pan, P. H. Shih, M. H. Wu, and J. H. Lin, “A storage assignment heuristic method based on genetic algorithm for a pick-and-pass warehousing system,” *Comput Ind Eng*, vol. 81, pp. 1–13, Mar. 2015, doi: 10.1016/j.cie.2014.12.010.

[32] H. Zhang *et al.*, “Layout design for intelligent warehouse by evolution with fitness approximation,” *IEEE Access*, vol. 7, pp. 166310–166317, 2019, doi: 10.1109/ACCESS.2019.2953486.

[33] T. Likhouzova and Y. Demianova, “Robot path optimization in warehouse management system,” *Evol Intell*, pp. 1–7, May 2021, doi: 10.1007/s12065-021-00614-w.

[34] H. Y. Lee and C. C. Murray, “Robotics in order picking: evaluating warehouse layouts for pick, place, and transport vehicle routing systems,” *Int J Prod Res*, vol. 57, no. 18, pp. 5821–5841, 2019, doi: 10.1080/00207543.2018.1552031.

[35] Department of Economic and Social Affairs, “Central Product Classification (CPC) Version 2.1,” New York, 2015. Accessed: Jun. 21, 2021. [Online]. Available: https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/CPCv2.1_complete(PDF)_English.pdf

[36] K. Buza, A. Buza, and P. B. Kis, “Towards better modeling of supermarkets,” in *ICCC-CONTI 2010 - IEEE International Joint Conferences on Computational Cybernetics and Technical Informatics, Proceedings*, 2010, pp. 499–503. doi: 10.1109/ICCCYB.2010.5491220.

[37] J. Han, M. Kamber, F. Berzal, and N. Marín, “Data Mining: Concepts and Techniques,” vol. 500, 2001.

[38] K. Garg, “Mining Efficient Association Rules Through Apriori Algorithm Using Attributes and Comparative Analysis of Various Association Rule Algorithms,” *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 6, p. 2277, 2013, Accessed: Mar. 21, 2022. [Online]. Available: www.ijarcsse.com

[39] M. J. Zaki and M. Ogihara, “Theoretical Foundations of Association Rules,” 2004, Accessed: Jan. 30, 2022. [Online]. Available: https://www.researchgate.net/publication/2921443

[40] F. Liang, H. Li, W. Zhang, and C. Zhang, “An Improved Distance Metric Clustering Algorithm for Association Rules,” *J Phys Conf Ser*, vol. 1284, no. 1, p. 012030, Aug. 2019, doi: 10.1088/1742-6596/1284/1/012030.

[41] N. Hussein, A. Alashqur, and B. Sowan, “Using the interestingness measure lift to generate association rules,” *Journal of Advanced Computer Science & Technology*, vol. 4, no. 1, pp. 156–162, 2015, doi: 10.14419/jacst.v4i1.4398.

[42] R. Agrawal and R. Srikant, “Fast Algorithms for Mining Association Rules”.

[43] C. Aflori and M. Craus, “Grid implementation of the Apriori algorithm,” *Advances in Engineering Software*, vol. 38, no. 5, pp. 295–300, 2007, doi: 10.1016/J.ADVENGSOFT.2006.08.011.

[44] F. J. M. Arboleda, G. P. Ortega, and J. A. G. Luna, “Temporal Visual Profiling of Market Basket Analysis,” *IAENG Int J Comput Sci*, vol. 49, no. 2, 2022.

[45] J. Leskovec, A. Rajaraman, and J. D. Ullman, “Clustering,” *Mining of Massive Datasets*, pp. 228–266, Dec. 2014, doi: 10.1017/CBO9781139924801.008.

[46] S. Theodoridis and K. Koutroumbas, “Chapter 13 - Clustering Algorithms II: Hierarchical Algorithms,” in *Pattern Recognition (Fourth Edition)*, Fourth Edition., S. Theodoridis and K. Koutroumbas, Eds. Boston: Academic Press, 2009, pp. 653–700. doi: https://doi.org/10.1016/B978-1-59749-272-0.50015-3.

[47] I. Kononenko and M. Kukar, “Cluster Analysis,” *Machine Learning and Data Mining*, pp. 321–358, Jan. 2007, doi: 10.1533/9780857099440.321.

[48] The Pennsylvania State University, “Hierarchical Clustering.” https://online.stat.psu.edu/stat555/node/85/ (accessed Mar. 27, 2022).

[49] “Agglomerative Hierarchical Clustering,” *The Pennsylvania State University*. https://online.stat.psu.edu/stat505/lesson/14/14.4 (accessed Mar. 28, 2022).

[50] “Hierarchical Clustering.” https://www.solver.com/xlminer/help/hierarchical-clustering-intro (accessed May 01, 2022).

[51] C. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[52] “Groceries Market Basket Dataset | Kaggle.” https://www.kaggle.com/datasets/irfanasrullah/groceries (accessed Oct. 14, 2022).

[53] A. F. Arboleda and F. Moreno, “Análisis y oportunidades para el diseño de supermercados basado en reglas de asociación,” in *Computación para el Desarrollo – XIV Congreso*, 2021, pp. 57–67.