

Examen Parcial No.2

Lea con atención las instrucciones que se le presentan. De manera individual, desarrolle las siguientes actividades. Tiene 90 min pura responder, utilice su criterio y conocimientos desarrollados durante el curso para ejecutar cada actividad. Al finalizar el examen, deberá enviar por la plataforma Moodle un zip que contenga los elementos siguientes:

1. Word con respuestas de Serie 1
2. Script de R con el código generado (no olvide comentar su procedimiento)
3. Presentación de PowerPoint requerida.

Serie 1 (25 puntos)

1. Desarrolle con sus palabras, ¿cuál es el objetivo de realizar clustering?

El Clustering es un tipo de Machine Learning que busca crear relaciones entre grupos de variables. Se aplica un aprendizaje no supervisado, ya que no requiere de datos previamente etiquetados y la relación que encuentra entre las variables se basa en las propias observaciones. Su objetivo principal es organizar la información de manera que las observaciones entre los grupos sean lo más similares entre si tanto como sea posible.

2. ¿Cuáles son los 3 tipos de algoritmos que existen en machine learning?

- Aprendizaje supervisado: Utiliza datos existentes clasificados a partir de etiquetas. Posee un histórico de cierto tipo de patrones (Ej. Identificador de SPAM).
- Aprendizaje no supervisado: No hay conocimiento previo. Existe un caos en los datos, busca identificar patrones y agruparlos. (Ej. Marketing y campañas de redes sociales).
- Aprendizaje por esfuerzo: Aprende a partir de la propia experiencia. Busca que sea capaz de tomar la mejor decisión a partir de prueba y error. (Ej. Reconocimiento facial).

3. Desarrolle con sus palabras, ¿cuál es el objetivo de realizar reglas de asociación?

El objetivo de realizar reglas de asociación es encontrar la cantidad de transacciones que contienen un conjunto En específico de artículos u objetos. Es decir, obtener la probabilidad de que un artículo sea obtenido dado "una condición". Usado generalmente sobre las compras de clientes, nos sirve para identificar ocurrencias de productos al hacer ciertas compras.

4. ¿Qué es un itemset?

Es un conjunto de artículos. Se dice que k-itemset es un set que contine k artículos. Por ejemplo
Conjunto { Cerveza, Pañales }

5. ¿Cuál es la diferencia entre el soporte y la confianza?

El soporte indica la cantidad de subconjuntos (itemset) se encuentra entre todas las transacciones, es decir, una probabilidad absoluta. Mientras que la confianza mide la frecuencia de aparición de los ítems en Y en transacciones que contienen X, en otras palabras, es una probabilidad condicional.

Serie 2 (50 puntos)

En la siguiente serie deberá utilizar las fuentes de datos indicadas para analizar la información usando R.

La empresa “Los Intensos” se dedica a la comercialización de productos de primera necesidad. Poseen una base de datos que contiene la información de sus clientes principales y sus gustos de consumo. Sin embargo, en los últimos años la gerencia de tecnología ha tenido los datos para uso exclusivo del departamento y no se ha aprovechado la información almacenada.

Se ha incorporado a la institución un nuevo Director General, con enfoque en datos y lo ha contratado a usted cómo analista de información para utilizar explorar los datos y encontrar estrategias para maximizar sus ventas. Por lo que le solicita los siguiente:

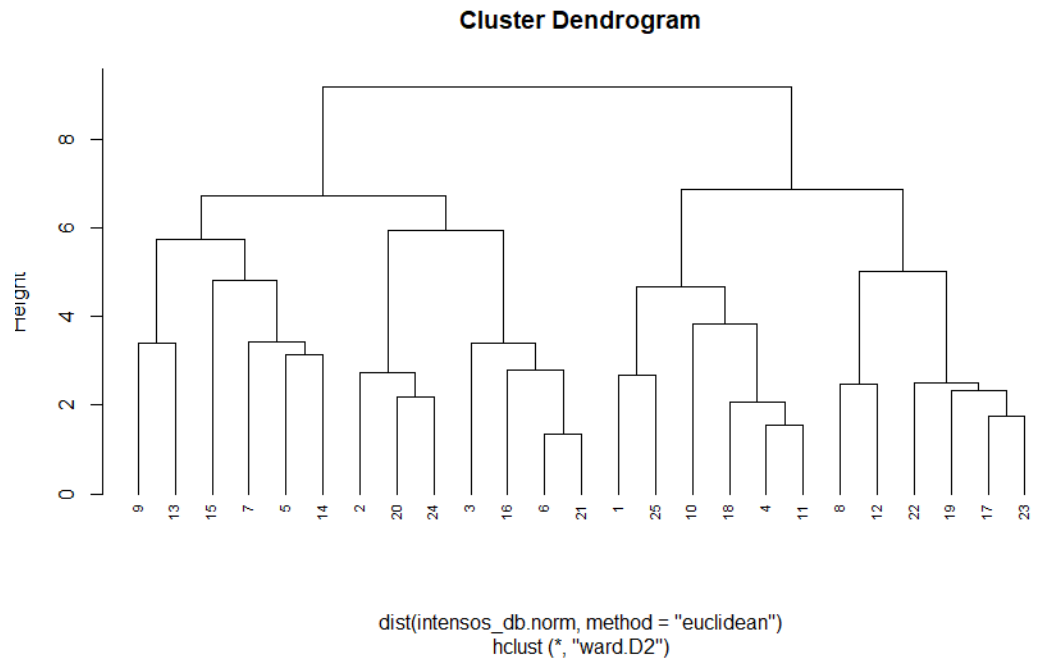
- a. De la base de datos llamada “intensos_db”
 - i. Genere la estadística General de los datos.

Cliente	Carne	Agua	Leche	Nueces	Cereal
Length:25	Min. : 0.4296	Min. : 0.09941	Min. : 0.06377	Min. : 0.8822	Min. : 0.07865
Class :character	1st Qu.: 3.1353	1st Qu.: 1.79001	1st Qu.: 0.44809	1st Qu.: 5.1127	1st Qu.: 0.64334
Mode :character	Median : 5.8138	Median : 3.45259	Median : 1.14183	Median : 7.8509	Median : 1.37780
	Mean : 5.8810	Mean : 4.52295	Mean : 1.42741	Mean : 10.0045	Mean : 1.81670
	3rd Qu.: 7.9707	3rd Qu.: 7.22635	3rd Qu.: 2.00900	3rd Qu.: 15.3853	3rd Qu.: 2.41599
	Max. : 16.7970	Max. : 13.00155	Max. : 4.21371	Max. : 24.1372	Max. : 8.21268
Cerveza	Jugos	Frutas	Vegetales		
Min. : 0.4668	Min. : 0.3288	Min. : 0.1010	Min. : 0.04992		
1st Qu.: 7.3672	1st Qu.: 0.5120	1st Qu.: 0.8314	1st Qu.: 1.36131		
Median : 12.1699	Median : 1.7803	Median : 1.6692	Median : 1.98572		
Mean : 15.5882	Mean : 2.1041	Mean : 2.0262	Mean : 2.22579		
3rd Qu.: 24.9895	3rd Qu.: 3.7010	3rd Qu.: 2.9417	3rd Qu.: 2.44335		
Max. : 42.9466	Max. : 4.5569	Max. : 4.9916	Max. : 6.68434		

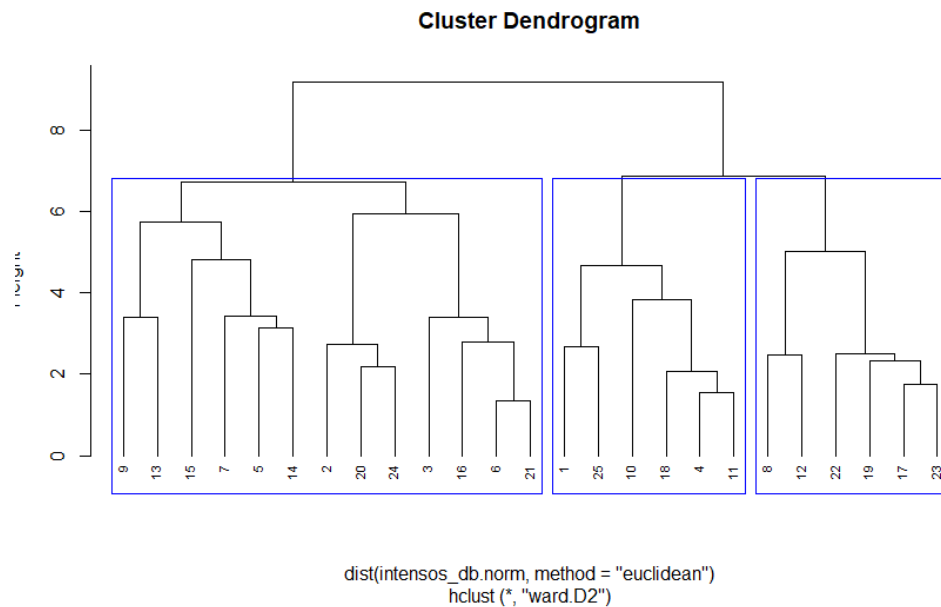
- b. Debe de realizar un script que permita por medio del algoritmo K-Means realizar los cluster que mejor agrupen la información dada.

c. Debe de crear y adjuntar los siguientes plots:

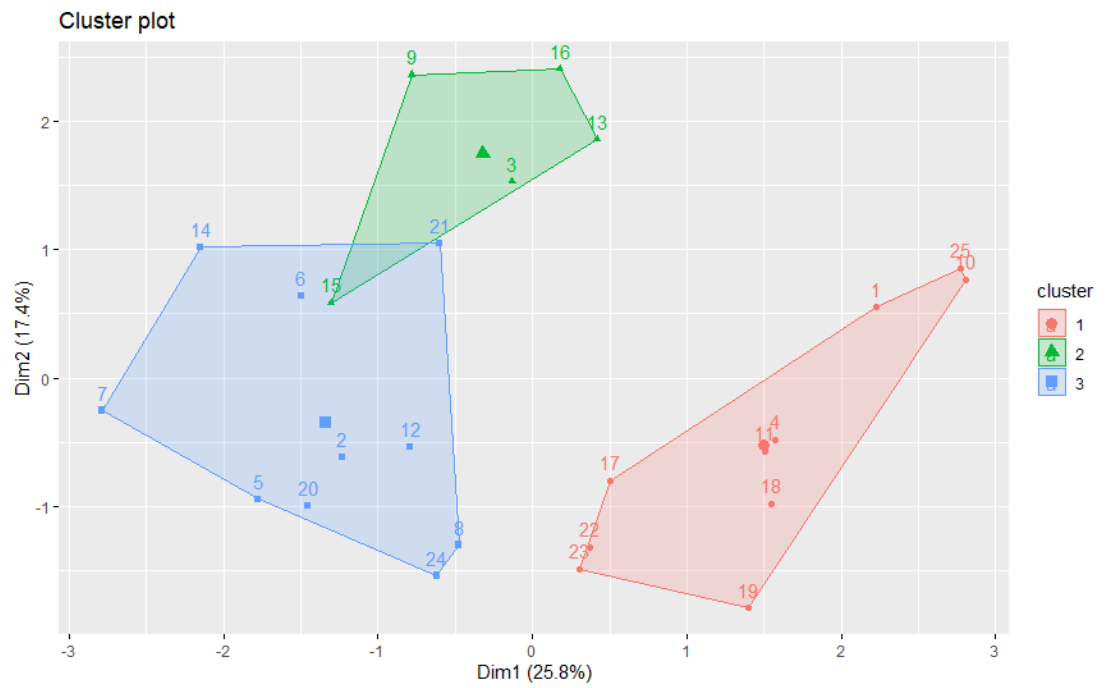
i. Un plot con el dendrograma del modelo jerárquico



ii. Un plot con los diferentes grupos clusters definidos (cuttree)



iii. Un plot utilizando factoextra para representar los k-means



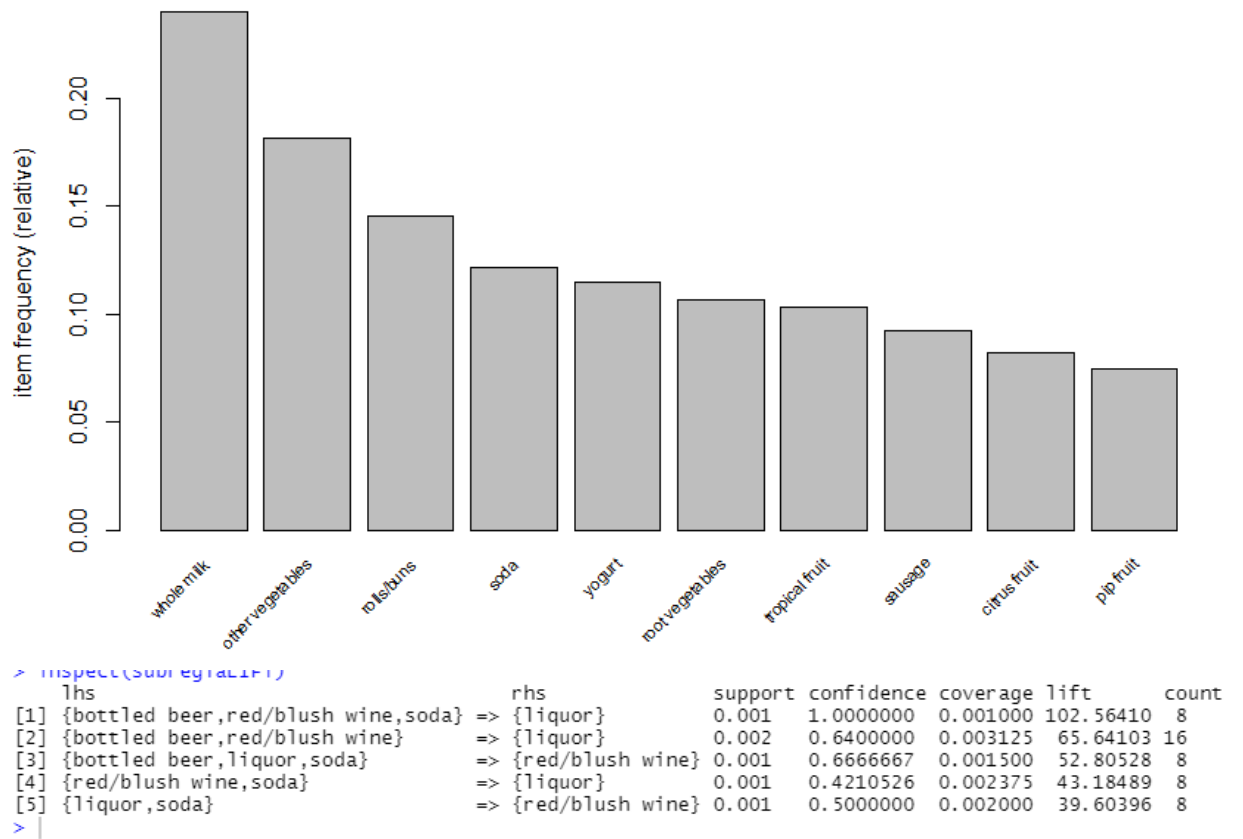
d. Debe de responder lo siguiente:

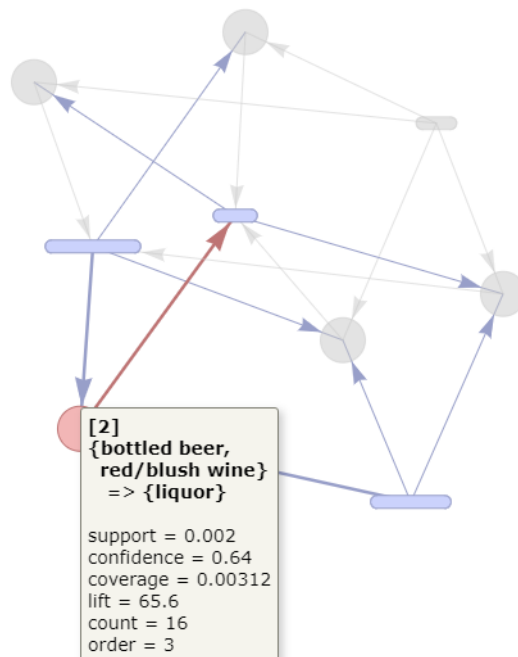
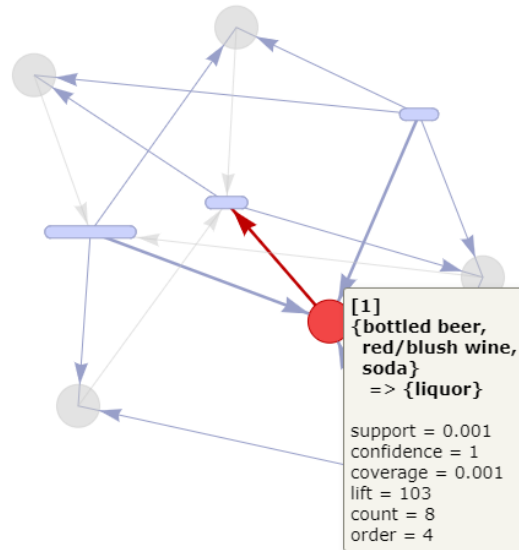
i. ¿Cuáles serían las etiquetas para cada cluster?

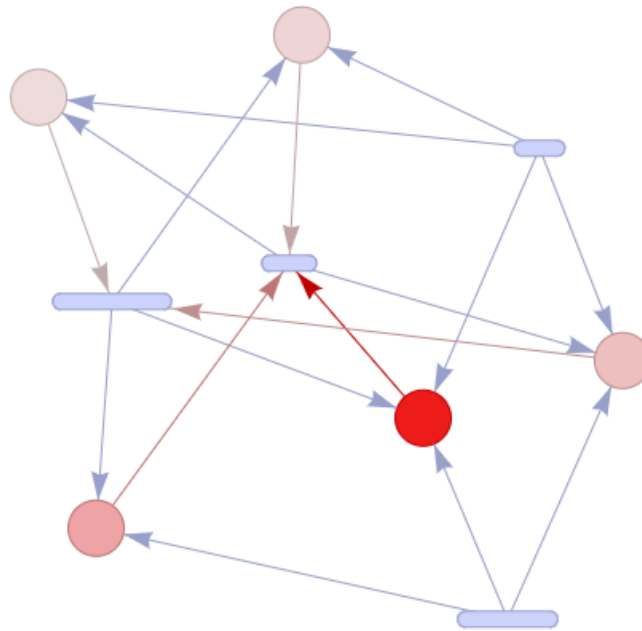
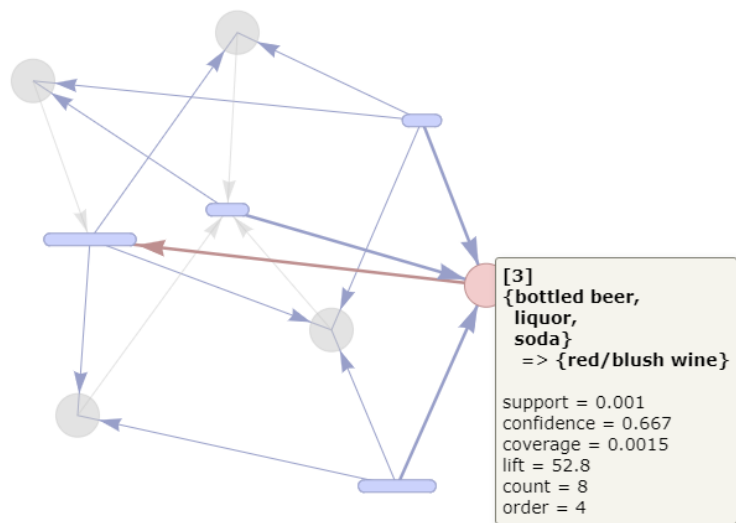
El número óptimo de clústers es de 3, ya que representa una diferencia clara, por lo tanto, se tienen 3 etiquetas. Estas etiquetas representarían los perfiles de los compradores, por ejemplo: están los que están más interesados en comprar carne, aquellos que prefieren frutas y vegetales, y por último los que prefieren otros (como cereales).

Teniendo claros los grupos por los que elaborará su estrategia de segmentación y ventas a clientes, establezca mediante el algoritmo a priori, las reglas de asociación de los productos que se comercializan con el objetivo de generar ofertas y promociones. Para ello deberá:

- e. Realizar un script que permita por medio del algoritmo a priori, generar las reglas de asociación del dataset (intensos_pdr).
- f. Por la dispersión de productos, debe utilizar un support = 0.001 y un confidence de 0.3
- g. Debe establecer que el mínimo de artículos para la regla será de 2
- h. Cree el plot que muestre las distintas reglas encontradas.







Serie 3 (25 puntos)

En una presentación de PowerPoint, desarrolle los siguientes elementos:

- i. Dendograma de clusters generados
- j. Teniendo las reglas creadas, analice las mismas haciendo uso de la métrica LIFT para seleccionar las reglas mayor relevancia, sin olvidar el uso de support y confidence.
- k. Seleccione la regla de asociación que evalúe cómo más relevantes y establezca una estrategia de venta para ella, enfocado a un cluster específico (según su análisis de estadística de consumo y preferencia del cluster).