

Agrupamiento y asociación

Indicación: deberá utilizar las fuentes de datos indicadas para analizar la información usando R.

Para cada uno de los ejercicios:

- Suba el archivo .R indicado
- Adjunte el PLOT como imagen en esta hoja de trabajo
- Responda las preguntas indicadas

1. Ejercicio # 1

- a. Cree un nuevo script .R llamado “Clustering Repuestos”
- b. Cree una conexión a la base de datos de COVID y cree una vista que lea la data en R usando el siguiente script:

```
select  cat.Nombre,  
        sum(p.Precio*cd.Cantidad) as TotalPorParte,  
        sum(cd.Cantidad) as TotalPartesCotizadas,  
        avg(cd.cantidad) as PromedioPartesCotizadas  
from    Cotizacion C left join  
        CotizacionDetalle CD on C.IDCotizacion = CD.IDCotizacion inner join  
        Partes P on P.ID_Parte = cd.ID_Parte INNER JOIN  
        Categoria cat on Cat.ID_Categoria = p.ID_Categoria  
group by cat.Nombre
```

- c. Debe realizar un script que permita, por medio de la técnica de agrupamiento de K-Means visto en clase, encontrar cuál es el comportamiento de los diferentes grupos de diferentes categorías, en esencia lo que se busca es poder clasificar a las categorías de partes en grupos que se puedan encontrar por medio del agrupamiento, encontrando cosas como, por ejemplo
 - i. Categorías más rentables
 - ii. Categorías mas con más partes cotizadas
- d. Debe crear y adjuntar los siguientes plots:
 - i. Un plot con los diferentes grupos (probar múltiples clusters)
 1. Dicho plot debe marcar los centros y elipses para cada cluster
 - ii. Un plot con el comportamiento de la minimización de distancias “tot.withinss” que muestre cómo se comporta para diferentes números de clusters
- e. Debe responder lo siguiente:
 - i. ¿Cuál es el mejor número de clusters? ¿Por qué?
 - ii. ¿Cuáles serían las etiquetas para cada cluster?

2. Ejercicio # 2

- a. Cree un nuevo script .R llamado “Asociación compras”
- b. Lea el archivo subido al portal llamado “Compras.csv”
 - i. Este archivo tiene el listado de algunas transacciones de las compras hechas en una tienda de electrónicos y se desea saber si existe alguna relación implícita entre la compra de algún conjunto de productos
- c. Utilizando las reglas de asociación vistas en clase con la librería “arules”, realice lo siguiente:
 - i. Cree un plot interactivo y uno de grafos usando “arulesViz” que tenga las reglas encontradas (adjunte imagen de cada plot)
 - ii. Cree un objeto que permita, por medio de la función “apriori”, encontrar si existe una relación entre grupos de ítems, con un soporte del 50% y una confianza del 90%
 - iii. ¿Existe alguna regla entre productos que tenga este nivel de confianza? ¿Cuáles?
 - iv. ¿Cuál es el nivel de lift de cada regla encontrada?