

Proyecto Final (Association Rules)

Profesor Marciano A. Moreno D.C.

27 de mayo de 2021

Introducción

Este notebook contiene las actividades a realizar como parte del proyecto final de la materia Data Mining del Tecnológico de Monterrey, Campus Estado de México para el ciclo escolar enero/junio de 2021.

Las actividades incluyen ejecución, personalización y codificación de estatutos del lenguaje R con diversas librerías de reglas de asociación, así como cálculos, análisis, investigación y exposición de conceptos relacionados.

La respuesta a todas y cada una de las actividades (implementación de código, notas y enlace a video) debe quedar registrada en el presente documento en formato R Markdown, exportada como documento PDF y publicada en el repositorio GitHub del alumno.

Fecha límite de entrega: sábado 5 de junio al cierre del día.

Recursos para la realización del proyecto

El proyecto podrá realizarse en un ambiente de desarrollo local o remoto, dependiendo de los recursos que disponga el alumno.

Recursos generales:

- Cuenta de usuario en GitHub.
- GitHub Desktop.
- Fork del repositorio principal de la clase
- Conectividad a Internet.
- Dispositivo de grabación de video.
- Cuenta para publicación de video (puede ser privado) como es el caso de YouTube.

Ambiente de desarrollo local:

- Procesador Intel i5, 8 GB RAM, 50 GB HD.
- R Studio Desktop Open Source Edition.
- Sistema operativo soportado por R Studio.
- Proyecto de R Studio con el repositorio GitHub del alumno.

Ambiente de desarrollo remoto:

- Cuenta gratuita en R Studio Cloud.
- Espacio de trabajo creado a partir del repositorio GitHub del alumno.

Paquetes de R

- `git2r`
- `arulesViz` (incluye `arules`)
- `dplyr`

Conceptos

Chunk de código: región del documento identificada por triple tilde invertida.

```
#Chunk de código
```

Chunk de código sin personalización: El alumno deberá ejecutar el chunk de código sin mayor personalización.

```
#Chunk de código sin personalización.  
print("Este es un chunk de código sin personalización.")
```

```
## [1] "Este es un chunk de código sin personalización."
```

Las actividades de codificación estarán identificadas con la clave **CODE-nn** previo al chunk y contarán con comentarios **#TODO:**.

CODE-01 Suma 2+3 Implementa la operación de adición de los enteros 2 y 3

```
#TODO: Implementa la operación de adición con los enteros: 2 y 3  
2 + 3
```

```
## [1] 5
```

Notas en markdown: El alumno escribirá las notas requeridas en las secciones anotadas con la clave **NOTE-nn**

Escribe la fecha actual (**NOTE-01**): 4 de junio de 2021.

Personalización del notebook

CODE-02 Datos generales Asigna tus datos a las siguientes variables:

```
NOMBRE_COMPLETO = "Joaquin Rios Corvera"  
DIGITOS_MATRICULA = 01375441
```

Instalación y carga de paquetes base

Instala los paquetes base de este notebook.

Los siguientes paquetes son requeridos para aspectos operativos de los ejercicios, no están relacionados con los temas de la materia. Si los paquetes no están instalados deberás retirar los comentarios de los siguientes estatutos y ejecutar el chunk. Coloca los comentarios nuevamente cuando hayas instalado los paquetes.

CODE-03 Instala `git2r`

```
#TODO: Retira los comentarios la primera vez para instalar los paquetes.  
#install.packages("git2r")
```

git2r es una interfaz para el sistema de control de versiones Git que usaremos para verificar que el número de commit del proyecto final sea el esperado. El valor esperado del commit será comunicado por separado.

```
library(git2r)
```

```
## Warning: package 'git2r' was built under R version 3.6.2
```

```
git2r::revparse_single(repository("."),"HEAD")
```

```
## [9c56d4a] 2021-06-05: Merge remote-tracking branch 'upstream/main' into main
```

Instalación y carga de paquetes para los ejercicios

Instala los paquetes requeridos para este notebook.

En caso que los no tengas será necesario que retires los comentarios y ejecutes los comandos de la siguiente celda.

Tip: Coloca nuevamente en comentario las líneas de abajo en cuanto hayas instalado los paquetes.

CODE-04 Instala arulesViz

```
#TODO: Retira los comentarios la primera vez para instalar los paquetes.  
#install.packages("arulesViz")
```

Carga la librería arulesViz (la cual carga automáticamente arules).

```
library("arulesViz")
```

```
## Warning: package 'arulesViz' was built under R version 3.6.2
```

```
## Loading required package: arules
```

```
## Warning: package 'arules' was built under R version 3.6.2
```

```
## Loading required package: Matrix
```

```
##
```

```
## Attaching package: 'arules'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## abbreviate, write
```

smallbasket: Fundamentos de Association Rules

Considera a `smallbasket` como la siguiente lista de transacciones, implementa el código necesario y responde a las solicitudes indicadas.

TID	items
Tr10	{beer, nuts, diapers}
Tr20	{beer, coffee, diapers}
Tr30	{beer, diapers, eggs}
Tr40	{nuts, eggs, milk}
Tr50	{nuts, coffee, diapers, eggs, milk}

Responde a las siguientes preguntas de smallbasket:

- Cantidad de transacciones (n) (**NOTE-02**): 5.
- Frecuencia absoluta del item {beer} (**NOTE-03**): 3.
- Frecuencia absoluta del item {nuts} (**NOTE-04**): 3.
- Frecuencia absoluta del itemset {beer, nuts} (**NOTE-05**): 1.
- Support del item {beer} (**NOTE-06**): 0.6.
- Support del itemset {beer, nuts} (**NOTE-07**): 0.2.

Implementaremos a `smallbasket` en R por medio de `list`.

```
smallbasket <- list(c("beer", "nuts", "diapers"),
  c("beer", "coffee", "diapers"),
  c("beer", "diapers", "eggs"),
  c("nuts", "eggs", "milk"),
  c("nuts", "coffee", "diapers", "eggs", "milk"))

names(smallbasket) <- paste("Tr", seq(from = 10, to = 50, by = 10), sep = "")
```

Crea un objeto `smalltx` de tipo `arules::transactions` a partir de la lista `smallbasket` el cual deberá tener la siguiente estructura lógica:

TID	beer	nuts	diapers	coffee	eggs	milk
Tr10	1	1	1	0	0	0
Tr20	1	0	1	1	0	0
Tr30	1	0	1	0	1	0
Tr40	0	1	0	0	1	1
Tr50	0	1	1	1	1	1

CODE-05 Crea una variable `smalltx` que sea un objeto `transactions` a partir de los datos en `smallbasket`.

```
#TODO: Modifica el código de abajo para transformar smallbaskets a un objeto de tipo transactions
smalltx <- as(smallbasket, "transactions")
```

Verifica que `smalltx` sea de tipo `transactions`.

```
class(smalltx)[1] == "transactions"
```

```
## [1] TRUE
```

Implementa en el chunk de abajo el código para visualizar la información general del objeto `smalltx` por medio de la función `summary()`, visualiza la salida **R Console** y responde en línea en el markdown a las preguntas de abajo.

```
summary(smalltx)
```

```
## transactions as itemMatrix in sparse format with
## 5 rows (elements/itemsets/transactions) and
## 6 columns (items) and a density of 0.5666667
##
## most frequent items:
##   diapers    beer    eggs    nuts  coffee (Other)
##         4         3         3         3         2         2
##
## element (itemset/transaction) length distribution:
## sizes
## 3 5
## 4 1
##
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      3.0      3.0      3.0      3.4      3.0      5.0
##
## includes extended item information - examples:
##   labels
## 1    beer
## 2  coffee
## 3 diapers
##
## includes extended transaction information - examples:
##   transactionID
## 1           Tr10
## 2           Tr20
## 3           Tr30
```

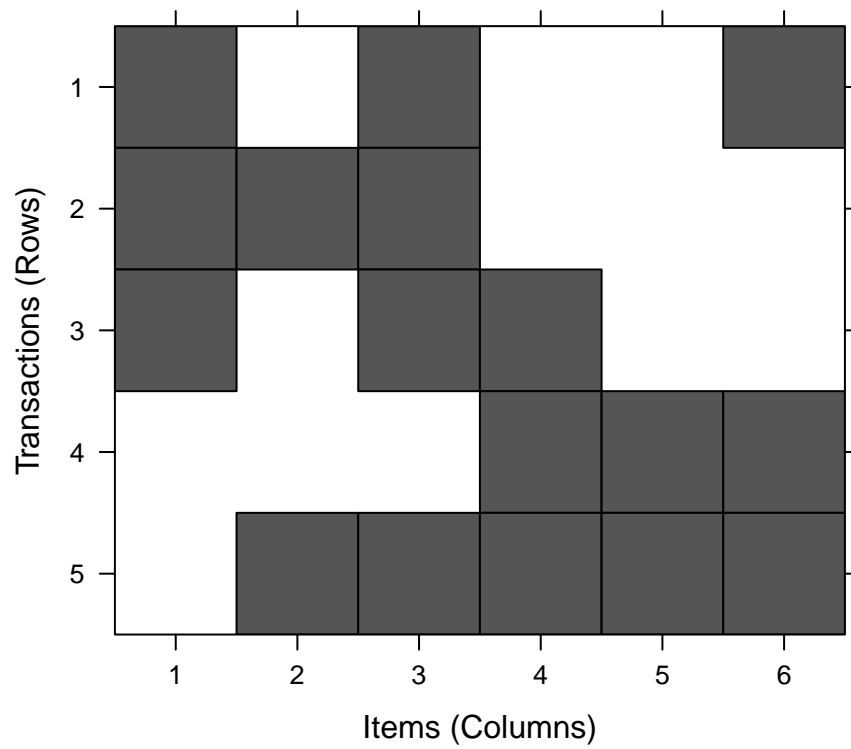
- Número de renglones (**NOTE-08**): 5.
- Número de columnas (**NOTE-09**): 6.
- Densidad de la matriz (**NOTE-10**): 0.5666667
- Media aritmética del número de elementos (itemsets) por transacción (**NOTE-11**): 3.4.

Visualiza la matriz de items y transacciones.

```
try(
  plot(smalltx, main = paste0("Elaborado por: ", NOMBRE_COMPLETO, " (", DIGITOS_MATRICULA, ")"))
)
```

```
## Warning in plot.itemMatrix(smalltx, main = paste0("Elaborado por: ",
## NOMBRE_COMPLETO, : Use image() instead of plot().
```

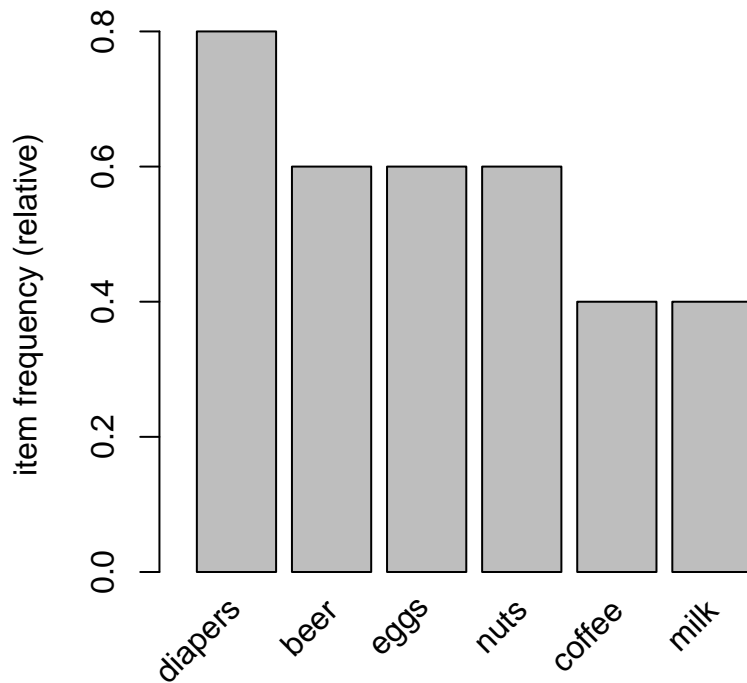
Elaborado por: Joaquin Rios Corvera (1375441)



Visualiza el gráfico de frecuencia de items.

```
try(  
  itemFrequencyPlot(smalltx, topN=10, main = paste0("Elaborado por: ", NOMBRE_COMPLETO, " (", DIGITOS_M  
)
```

Elaborado por: Joaquin Rios Corvera (1375441)



Invoca el algoritmo **apriori** de **arules** con las transacciones de **smalltx** y asignando el resultado a **smallrules**.

```
smallrules <- apriori(smalltx)
```

```
## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##          0.8    0.1    1 none FALSE                TRUE     5     0.1     1
## maxlen target  ext
##          10 rules TRUE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##       0.1 TRUE TRUE  FALSE TRUE     2    TRUE
##
## Absolute minimum support count: 0
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[6 item(s), 5 transaction(s)] done [0.00s].
## sorting and recoding items ... [6 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 done [0.00s].
## writing ... [46 rule(s)] done [0.00s].
## creating S4 object ... done [0.00s].
```

```
#parameter = list(support = 0.01, confidence = 0.01)
```

Menciona los valores por omisión de algoritmo apriori de la librería arules:

- support (mínimo) (**NOTE-12**): 0.1.
- confidence (mínimo) (**NOTE-13**): 0.8.
- items (máximo) (**NOTE-14**): 10.
- tiempo de verificación de subsets (máximo) (**NOTE-15**): 5 secs.

Invoca `summary()` con `smallrules`, consulta los resultados en **R Console** y responde a las preguntas de abajo.

Tip: Para mayor legibilidad emplea el comando **Show in new window**

```
summary(smallrules)
```

```
## set of 46 rules
##
## rule length distribution (lhs + rhs):sizes
##  1  2  3  4  5
##  1  4 18 18  5
##
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   3.000   3.500   3.478   4.000   5.000
##
## summary of quality measures:
##      support      confidence      coverage      lift
##   Min.   :0.2000   Min.   :0.8000   Min.   :0.2000   Min.   :1.000
##   1st Qu.:0.2000   1st Qu.:1.0000   1st Qu.:0.2000   1st Qu.:1.250
##   Median :0.2000   Median :1.0000   Median :0.2000   Median :1.667
##   Mean   :0.2478   Mean   :0.9957   Mean   :0.2522   Mean   :1.779
##   3rd Qu.:0.2000   3rd Qu.:1.0000   3rd Qu.:0.2000   3rd Qu.:2.500
##   Max.   :0.8000   Max.   :1.0000   Max.   :1.0000   Max.   :2.500
##      count
##   Min.   :1.000
##   1st Qu.:1.000
##   Median :1.000
##   Mean   :1.239
##   3rd Qu.:1.000
##   Max.   :4.000
##
## mining info:
##      data ntransactions support confidence
##   smalltx           5      0.1      0.8
```

- Cantidad de reglas producidas (**NOTE-16**): 46.
- Media aritmética de support (**NOTE-17**): 0.2478.
- Media aritmética de confidence (**NOTE-18**): 0.9957

Inspecciona `smallrules` y responde a las preguntas a continuación.


```
inspect(smallrules)
```

##	lhs	rhs	support	confidence	coverage
## [1]	{}	=> {diapers}	0.8	0.8	1.0
## [2]	{coffee}	=> {diapers}	0.4	1.0	0.4
## [3]	{milk}	=> {nuts}	0.4	1.0	0.4
## [4]	{milk}	=> {eggs}	0.4	1.0	0.4
## [5]	{beer}	=> {diapers}	0.6	1.0	0.6
## [6]	{coffee,milk}	=> {nuts}	0.2	1.0	0.2
## [7]	{coffee,nuts}	=> {milk}	0.2	1.0	0.2
## [8]	{coffee,milk}	=> {eggs}	0.2	1.0	0.2
## [9]	{coffee,eggs}	=> {milk}	0.2	1.0	0.2
## [10]	{coffee,milk}	=> {diapers}	0.2	1.0	0.2
## [11]	{diapers,milk}	=> {coffee}	0.2	1.0	0.2
## [12]	{beer,coffee}	=> {diapers}	0.2	1.0	0.2
## [13]	{coffee,nuts}	=> {eggs}	0.2	1.0	0.2
## [14]	{coffee,eggs}	=> {nuts}	0.2	1.0	0.2
## [15]	{coffee,nuts}	=> {diapers}	0.2	1.0	0.2
## [16]	{coffee,eggs}	=> {diapers}	0.2	1.0	0.2
## [17]	{milk,nuts}	=> {eggs}	0.4	1.0	0.4
## [18]	{eggs,milk}	=> {nuts}	0.4	1.0	0.4
## [19]	{eggs,nuts}	=> {milk}	0.4	1.0	0.4
## [20]	{diapers,milk}	=> {nuts}	0.2	1.0	0.2
## [21]	{diapers,milk}	=> {eggs}	0.2	1.0	0.2
## [22]	{beer,nuts}	=> {diapers}	0.2	1.0	0.2
## [23]	{beer,eggs}	=> {diapers}	0.2	1.0	0.2
## [24]	{coffee,milk,nuts}	=> {eggs}	0.2	1.0	0.2
## [25]	{coffee,eggs,milk}	=> {nuts}	0.2	1.0	0.2
## [26]	{coffee,eggs,nuts}	=> {milk}	0.2	1.0	0.2
## [27]	{coffee,milk,nuts}	=> {diapers}	0.2	1.0	0.2
## [28]	{coffee,diapers,milk}	=> {nuts}	0.2	1.0	0.2
## [29]	{coffee,diapers,nuts}	=> {milk}	0.2	1.0	0.2
## [30]	{diapers,milk,nuts}	=> {coffee}	0.2	1.0	0.2
## [31]	{coffee,eggs,milk}	=> {diapers}	0.2	1.0	0.2
## [32]	{coffee,diapers,milk}	=> {eggs}	0.2	1.0	0.2
## [33]	{coffee,diapers,eggs}	=> {milk}	0.2	1.0	0.2
## [34]	{diapers,eggs,milk}	=> {coffee}	0.2	1.0	0.2
## [35]	{coffee,eggs,nuts}	=> {diapers}	0.2	1.0	0.2
## [36]	{coffee,diapers,nuts}	=> {eggs}	0.2	1.0	0.2
## [37]	{coffee,diapers,eggs}	=> {nuts}	0.2	1.0	0.2
## [38]	{diapers,eggs,nuts}	=> {coffee}	0.2	1.0	0.2
## [39]	{diapers,milk,nuts}	=> {eggs}	0.2	1.0	0.2
## [40]	{diapers,eggs,milk}	=> {nuts}	0.2	1.0	0.2
## [41]	{diapers,eggs,nuts}	=> {milk}	0.2	1.0	0.2
## [42]	{coffee,eggs,milk,nuts}	=> {diapers}	0.2	1.0	0.2
## [43]	{coffee,diapers,milk,nuts}	=> {eggs}	0.2	1.0	0.2
## [44]	{coffee,diapers,eggs,milk}	=> {nuts}	0.2	1.0	0.2
## [45]	{coffee,diapers,eggs,nuts}	=> {milk}	0.2	1.0	0.2
## [46]	{diapers,eggs,milk,nuts}	=> {coffee}	0.2	1.0	0.2
##	lift	count			
## [1]	1.000000	4			
## [2]	1.250000	2			
## [3]	1.666667	2			

```

## [4] 1.666667 2
## [5] 1.250000 3
## [6] 1.666667 1
## [7] 2.500000 1
## [8] 1.666667 1
## [9] 2.500000 1
## [10] 1.250000 1
## [11] 2.500000 1
## [12] 1.250000 1
## [13] 1.666667 1
## [14] 1.666667 1
## [15] 1.250000 1
## [16] 1.250000 1
## [17] 1.666667 2
## [18] 1.666667 2
## [19] 2.500000 2
## [20] 1.666667 1
## [21] 1.666667 1
## [22] 1.250000 1
## [23] 1.250000 1
## [24] 1.666667 1
## [25] 1.666667 1
## [26] 2.500000 1
## [27] 1.250000 1
## [28] 1.666667 1
## [29] 2.500000 1
## [30] 2.500000 1
## [31] 1.250000 1
## [32] 1.666667 1
## [33] 2.500000 1
## [34] 2.500000 1
## [35] 1.250000 1
## [36] 1.666667 1
## [37] 1.666667 1
## [38] 2.500000 1
## [39] 1.666667 1
## [40] 1.666667 1
## [41] 2.500000 1
## [42] 1.250000 1
## [43] 1.666667 1
## [44] 1.666667 1
## [45] 2.500000 1
## [46] 2.500000 1

```

- ¿Cuál es la interpretación de la regla $\{\} \Rightarrow \{\text{diapers}\}$ en la que se aprecia la ausencia del LHS? (**NOTE-19**): El lift de 1 nos dice que no existe una relación entre el body y el head, es decir, la aparición del lado izquierdo no afecta la del lado derecho. Específicamente nos dice que sin importar los otros elementos de la transacción, diapers aparece lo suficiente para cumplir con la confidence establecida por nosotros (esto se puede ver en la tabla de frecuencias de arriba).
- Calcula el valor de la métrica support para la regla $\{\text{beer}\} \Rightarrow \{\text{nuts}\}$ (**NOTE-20**): $1/5 = 0.2$.
- Calcula el valor de la métrica confidence para la regla $\{\text{beer}\} \Rightarrow \{\text{nuts}\}$ (**NOTE-21**): $1/3 = 0.333$.
- ¿Por qué no aparece listada la regla $\{\text{beer}\} \Rightarrow \{\text{nuts}\}$? (**NOTE-22**): Mencionamos arriba que el confidence mínimo para el algoritmo es de 0.8, y esta regla no lo cumple al tener 0.333.

MSSD: El dataset de sesiones de streaming

The Music Streaming Sessions Dataset fue desarrollado por Spotify para promover la investigación en modelado de escucha por parte de usuarios, interacciones en streaming, recuperación de información musical (MIR, por sus siglas en inglés) y recomendaciones con base a sesiones secuenciales.

Analiza una extracción del dataset MSSD por medio de reglas de asociación.

Instala el paquete dplyr para llevar preparar los datos de MSSD de tal forma que puedan ser analizados por medio de arules.

CODE-06 Instala el paquete dplyr

```
#TODO: Retira los comentarios la primera vez para instalar los paquetes.  
#install.packages("dplyr")
```

Carga la librería dplyr.

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.6.2
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:arules':
```

```
##
```

```
## intersect, recode, setdiff, setequal, union
```

```
## The following object is masked from 'package:git2r':
```

```
##
```

```
## pull
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

Lee el dataset MSSD y verifica su estructura.

```
mssddf <- read.csv("/Users/joaquin/Desktop/mssd-log_mini.csv") #RStudio crasheaba al intentar leerlo de  
str(mssddf)
```

```
## 'data.frame': 167880 obs. of 21 variables:
```

```
## $ session_id : Factor w/ 10000 levels "0_00006f66-33e5-4de7-a324-2d18e439fc1e",
```

```
## $ session_position : int 1 2 3 4 5 6 7 8 9 10 ...
```

```
## $ session_length : int 20 20 20 20 20 20 20 20 20 20 ...
```

```
## $ track_id_clean : Factor w/ 50704 levels "t_00007fba-6bd3-449d-85dd-54d4aea397c2",
```

```
## $ skip_1 : Factor w/ 2 levels "false","true": 1 1 1 1 1 1 2 2 1 2 ...
```

```
## $ skip_2 : Factor w/ 2 levels "false","true": 1 1 1 1 1 1 2 2 2 2 ...
## $ skip_3 : Factor w/ 2 levels "false","true": 1 1 1 1 1 2 2 2 2 2 ...
## $ not_skipped : Factor w/ 2 levels "false","true": 2 2 2 2 2 1 1 1 1 1 ...
## $ context_switch : int 0 0 0 0 0 0 0 0 0 0 ...
## $ no_pause_before_play : int 0 1 1 1 1 1 1 1 1 1 ...
## $ short_pause_before_play : int 0 0 0 0 0 0 0 0 0 0 ...
## $ long_pause_before_play : int 0 0 0 0 0 0 0 0 0 0 ...
## $ hist_user_behavior_n_seekfwd : int 0 0 0 0 0 0 0 0 0 0 ...
## $ hist_user_behavior_n_seekback : int 0 0 0 0 0 0 0 0 0 0 ...
## $ hist_user_behavior_is_shuffle : Factor w/ 2 levels "false","true": 2 2 2 2 2 2 2 2 2 2 ...
## $ hour_of_day : int 16 16 16 16 16 16 16 16 16 16 ...
## $ date : Factor w/ 18 levels "2018-05-12","2018-05-18",...: 17 17 17 17 17 17 ...
## $ premium : Factor w/ 2 levels "false","true": 2 2 2 2 2 2 2 2 2 2 ...
## $ context_type : Factor w/ 6 levels "catalog","charts",...: 3 3 3 3 3 3 3 3 3 3 ..
## $ hist_user_behavior_reason_start: Factor w/ 9 levels "appload","backbtn",...: 8 8 8 8 8 8 5 5 5 5 .
## $ hist_user_behavior_reason_end : Factor w/ 7 levels "backbtn","clickrow",...: 7 7 7 7 7 4 4 4 4 4
```

Ispecciona la columna `session_id`.

```
summary(mssddf$session_id)
```

```
## 0_00006f66-33e5-4de7-a324-2d18e439fc1e 0_0000a72b-09ac-412f-b452-9b9e79bde8f
##                                20                                20
## 0_00010fc5-b79e-4cdf-bc4c-f140d0f99a3a 0_00016a3d-9076-4f67-918f-f29e3ce160dc
##                                20                                20
## 0_0002fbb5-29c2-4d54-9636-e6d15e77b58c 0_00030fe3-298d-4930-8534-fc5a9b82971b
##                                20                                20
## 0_0003ac37-c70a-47a6-b9dc-b8b163826082 0_0003e8a0-ea43-47fe-ac5a-8b502f4d38ef
##                                20                                20
## 0_00041f88-6d8a-49ca-9a03-d0cfdbbaf5e 0_00052b3c-a5f9-4b04-90ef-6378777d89d8
##                                20                                20
## 0_0005a7a1-196e-4e08-8928-39b9a3ca2e1a 0_000641b5-a2ce-4803-bdeb-499c548ba50d
##                                20                                20
## 0_0006bfe8-8afe-45c2-a423-00a348b775c4 0_0006f5f8-bd71-4d75-9cd2-6320f2e4df8e
##                                20                                20
## 0_0007cd53-3aa5-49ba-9473-319d5d7b2534 0_0008bd24-bbb0-4747-b777-7c97ce5ab3a9
##                                20                                20
## 0_0009367d-0072-4f19-8458-e291ddbbbd25 0_00094b63-0ef1-4cec-94e5-30879a3cce28
##                                20                                20
## 0_0009bdeb-b5b7-4073-9b01-b72a32c77cda 0_000a2236-9dae-4f60-b2a6-2420574a4cdc
##                                20                                20
## 0_000af7ac-0eb1-4f29-bc50-faf8155a0812 0_000bec52-cae6-4e1f-9518-18a4672ed8a3
##                                20                                20
## 0_000cd756-22f7-4551-8a43-ea8b88cfc1a9 0_000d307c-cfb4-4a65-9909-7b54d6b39cc8
##                                20                                20
## 0_000df85e-b4c8-40dc-82cf-25f483effe20 0_000e69be-6083-48e9-81c7-8e45ff1794df
##                                20                                20
## 0_000ef258-8f30-4351-b0d4-d61697b20fa3 0_000fccb3-2639-46b8-b766-81167a5bce01
##                                20                                20
## 0_000fd7ed-3e38-4a30-8a4e-41abc122c83a 0_000ff87d-b0de-47c7-8fea-b4a65ee2d089
##                                20                                20
## 0_00104e1b-6e8f-4931-a079-a29d7991eea6 0_0010721f-d313-4a67-989e-b32cc44665dd
##                                20                                20
```

```

## 0_001179c0-5bf5-4008-ab47-b10258ecf9f7 0_00119bb4-131f-4a2a-9cf0-ddff187993e5
##                                     20                                     20
## 0_00119c08-cfd8-4717-af77-4f22a1f10387 0_00122af2-9ce1-4da6-9251-1a99d5f11195
##                                     20                                     20
## 0_001254f5-ccf0-4a48-8484-e6703e3d1f63 0_00127e4f-400b-4284-8e2c-e9accbb9c54b
##                                     20                                     20
## 0_00130aaf-66be-4824-be3a-d41f8b8caafb 0_00137cc4-c737-4b01-9e72-1161f3723b72
##                                     20                                     20
## 0_0013efcc-bd57-4a73-b519-2c59fa42defb 0_00141184-1b81-4937-ad53-f05390e64a2b
##                                     20                                     20
## 0_001442ea-ca81-479e-9780-ae765c1b1bde 0_0014d04b-c05f-4e04-b677-9a1f5d182a6c
##                                     20                                     20
## 0_0014f558-7944-4060-8947-c3d863798ad8 0_0014f6a2-1c0b-4548-9ff2-f3b62fc0a96b
##                                     20                                     20
## 0_001544ae-d305-4178-9236-8181ced909f4 0_001666eb-6d08-43c2-883a-76c48d8fc34b
##                                     20                                     20
## 0_00189cd3-c830-43cd-9dc3-2241b54fcbb7 0_00189d78-d420-4b73-8f7f-836f795c7c06
##                                     20                                     20
## 0_001b1c67-69b3-4ec6-9c8c-e4f84fa9fba4 0_001bcdcd-5413-4f0d-bc9d-47ba460c63a4
##                                     20                                     20
## 0_001bd376-06fa-4fab-bbfb-a496f88e1cee 0_001dc5e3-6c41-431c-8c7b-6d073d9ea683
##                                     20                                     20
## 0_001de7c8-fde9-42a7-a3b1-39612d123762 0_001dfda4-9357-4136-9f56-4066f50e9d35
##                                     20                                     20
## 0_001e16a4-1d78-4d39-b191-df6b2cc97a02 0_001f6fca-d0fd-40ec-9b1c-0ec96e2fd06b
##                                     20                                     20
## 0_00202254-4b6b-45c2-9471-a5b26d5c751e 0_00207c2a-30b3-400f-99fb-52f77722ee47
##                                     20                                     20
## 0_0020bfcc-989f-4b5c-8082-c1121ef078cb 0_0021e01f-aa7d-4746-a4d6-10ec3741ec1f
##                                     20                                     20
## 0_0022f275-b84c-4bab-a0b7-ae5d4799b8fe 0_0023efe6-5658-4168-be67-45d3316cae0d
##                                     20                                     20
## 0_0024ad6a-c073-45cb-bf28-11032771e182 0_00250620-2a55-42f5-bf93-1b1b51b5fb31
##                                     20                                     20
## 0_0025179d-081b-4e18-a0f7-d2cd879e27b2 0_00259256-c8a2-4b6d-80bc-68b143d15a1c
##                                     20                                     20
## 0_00297b75-84de-4ea7-84df-bdd11e9fb665 0_0029d467-d1c5-4a83-8385-8ce0f8897b26
##                                     20                                     20
## 0_002ab260-d249-46fa-83f4-3d1c3cdb8616 0_002b2e36-21af-4055-897a-e8155d796584
##                                     20                                     20
## 0_002ba7ee-a3bc-4f40-aeef-39e07d7c39de 0_002ce345-c1a1-4b90-aac9-13b50b5c709f
##                                     20                                     20
## 0_0030107d-d09e-4284-aaa4-4a77d0d31869 0_003033f3-8953-4bcb-94df-62c227ab556e
##                                     20                                     20
## 0_00314960-6fbb-445e-8c89-5ec21a69f52c 0_0031f3c7-9869-4b6b-89c1-0b4bb04f8496
##                                     20                                     20
## 0_0032b965-74d5-49e9-8272-7b5948320e6e 0_00330c55-d1d7-427c-af10-52cb1b6b0361
##                                     20                                     20
## 0_003358d4-7930-4361-9830-48159a012f36 0_0033a562-767b-4694-9b2f-6c6817ec09f7
##                                     20                                     20
## 0_003553ab-c3e7-40d0-a9d3-2648581cd452 0_0035b35a-215a-49db-a49b-0b89ddf096bb
##                                     20                                     20
## 0_0037b3da-1968-4902-a79d-31f7bfb60c0e 0_003c00f6-4017-4c38-a7b6-96d76cd443bb
##                                     20                                     20
##

```

```
## 0_003c2b60-1527-48c0-8a89-928cfb08fc66 0_003e210a-ea37-4063-97b4-c6e055e3ea0e
##                                     20                                     20
## 0_003fd584-63a1-4411-9de5-08858a99a3f9 0_003fd817-60a2-4b5a-be9c-019988e2d0fd
##                                     20                                     20
## 0_003fdf83-8906-4cf2-a0d1-508b2c990bef 0_004068f2-38a7-4b8e-93b2-80e967c57c5f
##                                     20                                     20
## 0_004079d3-7ff5-4d30-8df9-74461cb37966 0_0040c1f3-634d-4a9d-9b59-28e388e7a69a
##                                     20                                     20
## 0_0040d206-cfdc-45d4-ae7a-d38656cf46c0 0_0040d4d8-a985-4b8e-8cf8-73cf6245029e
##                                     20                                     20
## 0_004178b5-8aa0-464e-ba4f-cb2bb1303d5e 0_0041c0a1-1811-4b22-98cc-24420789dde5
##                                     20                                     20
## 0_00435b28-1c47-4c84-b2ac-df59bd9745ad (Other)
##                                     20                                     165900
```

```
head(mssddf$session_id)
```

```
## [1] 0_00006f66-33e5-4de7-a324-2d18e439fc1e
## [2] 0_00006f66-33e5-4de7-a324-2d18e439fc1e
## [3] 0_00006f66-33e5-4de7-a324-2d18e439fc1e
## [4] 0_00006f66-33e5-4de7-a324-2d18e439fc1e
## [5] 0_00006f66-33e5-4de7-a324-2d18e439fc1e
## [6] 0_00006f66-33e5-4de7-a324-2d18e439fc1e
## 10000 Levels: 0_00006f66-33e5-4de7-a324-2d18e439fc1e ...
```

```
length(unique(mssddf$session_id))
```

```
## [1] 10000
```

Inspecciona la columna track_id_clean.

```
summary(mssddf$track_id_clean)
```

```
## t_bacf06d3-9185-4183-84ea-ff0db51475ce t_5718ab08-3a15-4d3f-9e63-42b2f6805e31
##                                     1427                                     915
## t_8c4d29b1-e0bf-464c-88f7-ac19240cbba0 t_a66ea088-b357-449a-8a1e-64dd0b8d6cb5
##                                     785                                     730
## t_77b02acb-1b1f-4b36-b8fc-2c3e01892b9a t_9a31436c-a57a-4d7a-9bbc-50f0deca33de
##                                     719                                     612
## t_0e3dec82-10b4-49f1-8c2e-cd19249f7d2c t_2fd6279a-06cd-4d5d-9e54-1912f51bb3c3
##                                     600                                     593
## t_7533e7ff-568e-448b-904e-fc0c3a9ec87e t_29a3895e-2c91-49a6-9383-6a71c597390d
##                                     576                                     548
## t_0ef6d2c8-faf9-4ebe-ac75-2bade8696a3d t_68888d19-7be6-4cdc-bcfb-be38431f79d7
##                                     536                                     528
## t_32270005-26a1-4763-8ba5-44fcc15f9914 t_a0fa9956-ffc3-41c9-a344-e247baa0e321
##                                     496                                     488
## t_8f9f2397-334b-48ee-9120-96d9ddf32f92 t_abe577df-bc3e-48ff-a13c-f2fd7a295ba6
##                                     484                                     484
## t_9229bbff-130d-4668-8e84-6b35c7b7cea0 t_312b3509-48c1-4928-a6e5-d3e7a51c216b
##                                     457                                     448
## t_81b9a202-281e-42a4-b068-e0c0f44e1e4d t_7e616ed3-115f-4a97-b90e-d128c392fae0
```

##	447	445
## t_1dba18ad-1ddb-4e97-983b-4af57e18d84a	t_f5205a27-c6b0-4c4d-b45f-312798372619	
##	437	422
## t_fb493d48-c98a-4101-8091-70d3a520d2c6	t_eb0063cc-948f-41d4-9dab-bf4b0e70d21e	
##	407	403
## t_338cb0f3-0035-418f-ab2d-3f9342815aaa	t_78b908e3-62db-4021-ac90-2ed146bcea8d	
##	383	380
## t_e8fa3463-7ef7-4ca5-8e65-2df979012c34	t_de860a61-bb57-4386-9d5a-6ba111c4c266	
##	354	341
## t_e9901b0e-dc9d-429c-a8a9-dde6ab877d54	t_4319026b-d2b5-478d-bf37-8f2873645c5c	
##	338	336
## t_8fd0c874-e115-4b75-8351-54acd7876edf	t_6dc02fe6-4d78-414a-8439-31240ee02270	
##	323	321
## t_bf17338d-830d-4890-b8e9-8ba41bb3349f	t_52fc9bcf-ce50-43fc-9498-c2c8421a33e7	
##	321	319
## t_ab55b9b5-3a1a-46a4-b58a-1c29136a08e2	t_b3db6567-f986-4b8f-bd7a-04057448ab3a	
##	313	308
## t_98e73634-8f33-4275-b544-3451553f431e	t_9a03f300-6504-4c7a-92b1-4cd88354ee74	
##	302	297
## t_297efa33-b070-4e5d-bb42-111237295b6c	t_a7aa673d-9c8b-4e6a-985e-ec91ed5727e9	
##	291	289
## t_e8308b79-41cf-4b31-b159-df093b592b41	t_b598e81d-c7dc-4bee-8c30-308e7b0cca2f	
##	288	283
## t_b193cd79-b6cb-4514-8c5b-ead28ac8dd30	t_58b65d28-3a54-467e-bf44-511aa89f44d4	
##	281	279
## t_b362a386-e30e-46be-bb44-af6efd02260f	t_fc5df5ba-5396-49a7-8b29-35d0d28249e0	
##	278	277
## t_992664c2-0754-4eb1-b0c3-0a6aa87697f5	t_f5cc5e37-ae7f-4788-a8e2-bbffb66eae62	
##	273	268
## t_2a55e695-7894-42aa-b920-52d1cbaf164c	t_6d21a6b7-bdac-46cc-bc45-9c731cbc7233	
##	263	262
## t_e906abc1-5808-4517-8ee4-2790c42c8ca4	t_6a40d9f0-3c44-4e11-86cd-b9a55fe2ce0a	
##	260	237
## t_3708b5d6-9d52-4470-9484-fff045aba7b6	t_42d15744-c948-489e-a92b-9e664c170b57	
##	234	232
## t_4ffd14f1-5406-4cfa-a9ed-a91a0f49b5bd	t_c0de25db-8c01-4122-9991-27132e852a18	
##	226	224
## t_bfe797e1-ddb7-4018-a6ba-ca05886533e7	t_f3c3c11b-76c3-401e-afbc-0182f37629bc	
##	218	216
## t_204562f8-e089-4305-be51-19913efdfbc6	t_61a0db14-98a8-4ad1-a770-a3ea9baba121	
##	214	212
## t_35bfb822-c716-491d-93fa-7e8dc0c12d07	t_d0c22c8c-7969-4d17-bf0d-d4fa3c04a768	
##	203	196
## t_f9e716a4-2403-4c9c-b84e-90e140b40750	t_fec4d8aa-7ff4-418b-b3d8-835ca0f7672e	
##	196	192
## t_f35eb373-15d1-4faa-9ca0-d0da975deebb	t_bfe9ae70-8add-43f6-903c-faf793b2d9c1	
##	179	176
## t_668f90a4-3d87-4044-a874-2388b4fac8b0	t_0f123593-5cb5-42ec-b4db-fb291a2e5eb9	
##	168	167
## t_ccaa489d-2ce8-4267-ac9e-99c5a4bd5f82	t_e78182e6-b5d6-427c-a269-9eaffff6165a	
##	163	162
## t_46ca5054-e280-4f02-a874-e362e343f479	t_8107cef1-6f3b-460d-b70a-b3f440b4b3f6	
##	160	158
## t_2e6b3753-5b16-4aa2-97f8-c77a81377139	t_b06cf9b1-8198-4b9f-ad64-381f500afb58	

```
##                                156                                155
## t_aae74c4e-7ac0-4ffc-8560-88303a918c3f t_ce600660-bdd3-4989-9436-54c7171e8eb9
##                                150                                149
## t_00750953-913d-4950-94a2-285b662c32f3 t_9a008295-d9a4-45f6-a179-7288879cf497
##                                148                                148
## t_c62a6914-7fdb-4165-8e3e-77299df1372c t_68f2d9f5-bb98-4dbe-8144-c83bf8cbff41
##                                147                                144
## t_244d8a7e-a290-424d-9b2e-05b06426914d t_e2dda38d-f826-481b-b49a-a1d9051f5c6d
##                                142                                142
## t_40c8e69e-f5d9-48ff-a9de-e62fcb78d3b7 t_92c7a02c-205b-4ca5-87bf-a244e6baa8cd
##                                141                                141
## t_adc08146-4685-4008-84c4-d37f28de0731 t_561a6648-8fe2-48fc-8a74-1cf13abd2d3b
##                                141                                140
## t_2a58f6bc-feb5-47f2-89bc-d1c28c14eff1 t_5d738bbe-951a-40a1-8680-fa56550bd977
##                                132                                132
## t_6af5de9f-d439-4da0-8c73-f7ba4ef3936b t_7c8d4b4b-af60-4d10-82c5-a048e3bed613
##                                132                                132
## t_c42b5e2e-a149-4063-b941-7cfdad1e6882 t_1d8eb515-169d-4eb6-baa3-0b519d499f8f
##                                132                                128
## t_363a6e62-e476-44d0-8a4c-575c62b2e45f t_36845b5e-e8af-42f3-b619-f85c9a7ddb21
##                                128                                128
## t_f9d9ad34-8301-45c9-9dad-980281c038c4 t_bc2844e7-dc32-4e72-ae18-eef4a32d9d68
##                                128                                127
## t_3c70d8ac-b601-4f8e-be57-cfdb7f119183 t_8d0c1ac0-afa4-400b-9560-38b850e966d2
##                                126                                126
## t_a004df37-97e0-4c20-999e-2844ddf9e733                                (Other)
##                                126                                137811
```

```
head(mssddf$track_id_clean)
```

```
## [1] t_0479f24c-27d2-46d6-a00c-7ec928f2b539
## [2] t_9099cd7b-c238-47b7-9381-f23f2c1d1043
## [3] t_fc5df5ba-5396-49a7-8b29-35d0d28249e0
## [4] t_23cff8d6-d874-4b20-83dc-94e450e8aa20
## [5] t_64f3743c-f624-46bb-a579-0f3f9a07a123
## [6] t_c815228b-3212-4f9e-9d4f-9cb19b248184
## 50704 Levels: t_00007fba-6bd3-449d-85dd-54d4aea397c2 ...
```

```
length(unique(mssddf$track_id_clean))
```

```
## [1] 50704
```

Consulta el paper The Music Streaming Sessions Dataset y responde a las siguientes preguntas:

- ¿Cuál es el propósito de `session id`? (**NOTE-23**): Es el ID de las sesiones, las cuales se refieren a una “listening session”, o el conjunto de todas las canciones que se escucharon de manera consecutiva por algún usuario.
- ¿Cuál es el propósito de `track id`? (**NOTE-24**): Es el ID de alguna canción en específico.
- ¿Consideras que las otras columnas del dataset son aplicables para análisis por medio de reglas de asociación, justifica tu respuesta? (**NOTE-25**): Claro, depende de qué se está buscando, creo que cada columna tiene algún uso, ya sea para ver las canciones que se saltan (`skip#`, `reason_end`), si las canciones se escuchan deliberadamente o simplemente por estar en una playlist (`shuffle`), o cualquier otra relación que se pueda formar entre ellas.

Inspecciona la estructura de y, el dataset preparado.

```
x <- mssddf %>% select(session_id, track_id_clean) %>% group_by(session_id)
y <- as.data.frame(x)
str(y)
```

```
## 'data.frame': 167880 obs. of 2 variables:
## $ session_id : Factor w/ 10000 levels "0_00006f66-33e5-4de7-a324-2d18e439fc1e",...: 1 1 1 1 1 1 1 1 1 1
## $ track_id_clean: Factor w/ 50704 levels "t_00007fba-6bd3-449d-85dd-54d4aea397c2",...: 891 28795 499
```

Descarga la librería dplyr

```
#Descargamos dplyr porque enmascara otras funciones y no es requerida en lo sucesivo
detach(package:dplyr)
```

Genera la representación de transacciones para arules del dataset MSSD y visualiza la salida **R Console**.

```
mssdtx <- as(split(y[, "track_id_clean"], y["session_id"]), "transactions")
```

```
## Warning in asMethod(object): removing duplicated items in transactions
```

```
#Trabajaremos con una muestra del 80% de las transacciones
numSessions <- round(nrow(mssdtx) * 0.8)
set.seed(DIGITOS_MATRICULA)
mssdtx <- sample(mssdtx, numSessions)
summary(mssdtx)
```

```
## transactions as itemMatrix in sparse format with
## 8000 rows (elements/itemsets/transactions) and
## 50704 columns (items) and a density of 0.0002874502
##
## most frequent items:
## t_bacf06d3-9185-4183-84ea-ff0db51475ce t_5718ab08-3a15-4d3f-9e63-42b2f6805e31
##                                     883                                     587
## t_8c4d29b1-e0bf-464c-88f7-ac19240cbba0 t_77b02acb-1b1f-4b36-b8fc-2c3e01892b9a
##                                     517                                     499
## t_a66ea088-b357-449a-8a1e-64dd0b8d6cb5                                     (Other)
##                                     494                                     113619
##
## element (itemset/transaction) length distribution:
## sizes
##      2      3      4      5      6      7      8      9     10     11     12     13     14     15     16     17
##     21     21     40     91    133    171    244    325    678    594    551    451    489    439    504    379
##     18     19     20
##    799    502   1568
##
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      2.00   11.00   15.00   14.57   19.00   20.00
##
## includes extended item information - examples:
##                                     labels
```

```
## 1 t_00007fba-6bd3-449d-85dd-54d4aea397c2
## 2 t_0000dc06-0c00-4a09-9dc6-3bdad9c6f0e8
## 3 t_00020dc1-1b82-43e9-8327-77b074bdf626
##
## includes extended transaction information - examples:
##                               transactionID
## 3881 0_059c613b-81e4-44b0-b5cf-9eefd49c0d75
## 5600 0_082ab48a-793d-4697-a012-de209fd67c3a
## 401  0_008e2d4a-44d4-4760-8e74-beb05837d487
```

Responde a las siguientes preguntas:

- ¿Cuál es el ID y frecuencia absoluta de la sesión más frecuente? (**NOTE-26**): t_bacf06d3-9185-4183-84ea-ff0db51475ce, frecuencia 883.
- ¿Cuántos items tiene el itemset con mayor número de ocurrencias en la lista element length distribution? (**NOTE-27**): 1568
- En el contexto de este dataset con información de sesiones de streaming, ¿cuál es la interpretación de la media aritmética? (**NOTE-28**): Se refiere a que la sesión promedio se conforma por 14.57 tracks.

Ejecuta el algoritmo apriori con los parámetros por omisión.

```
rules <- apriori(mssdtx)
```

```
## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##           0.8   0.1   1 none FALSE                TRUE     5     0.1     1
## maxlen target  ext
##          10 rules TRUE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##      0.1 TRUE TRUE  FALSE TRUE     2    TRUE
##
## Absolute minimum support count: 800
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[43601 item(s), 8000 transaction(s)] done [0.08s].
## sorting and recoding items ... [1 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 done [0.00s].
## writing ... [0 rule(s)] done [0.00s].
## creating S4 object ... done [0.01s].
```

```
summary(rules)
```

```
## set of 0 rules
```

- ¿Qué factores están asociados con que el algoritmo `apriori` con parámetros por omisión no haya producido ninguna regla de asociación, justifica tu respuesta? (**NOTE-29**): Como se puede observar en el DF, el valor default de confidence mínima es de 0.8, lo cual es demasiado alto para este set de datos, debido a su baja densidad de datos.

- ¿Qué riesgos identificas con la asignación de valores muy bajos en los parámetros de **support** y **confidence** al invocar el algoritmo **apriori**? Con una cantidad de datos tan grande como esta, valores muy bajos en los parámetros puede causar que se generen muchas más reglas de las necesarias, lo cual solamente vuelve al programa más lento sin dar beneficio alguno, ya que niveles tan bajos de support y confidence pueden generar reglas que no tengan tanta relevancia.

CODE-07 Invoca **apriori** Configura la invocación a **apriori** para producir al menos 220 reglas de asociación para ello puedes considerar el uso de los parámetros **support** y **confidence** como se muestra a continuación

```
rules <- apriori(mssdtx, parameter = list(support = SUPP, confidence = CONF))
```

```
#TODO: Configura las variables SUPP y CONF para generar al menos 250 reglas de asociación
SUPP = 0.02
CONF = 0.9
rules <- apriori(mssdtx, parameter = list(support = SUPP, confidence = CONF))
```

```
## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##      0.9      0.1      1 none FALSE          TRUE      5      0.02      1
## maxlen target  ext
##      10 rules TRUE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##      0.1 TRUE TRUE  FALSE TRUE      2      TRUE
##
## Absolute minimum support count: 160
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[43601 item(s), 8000 transaction(s)] done [0.06s].
## sorting and recoding items ... [56 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 6 7 done [0.00s].
## writing ... [883 rule(s)] done [0.00s].
## creating S4 object ... done [0.01s].
```

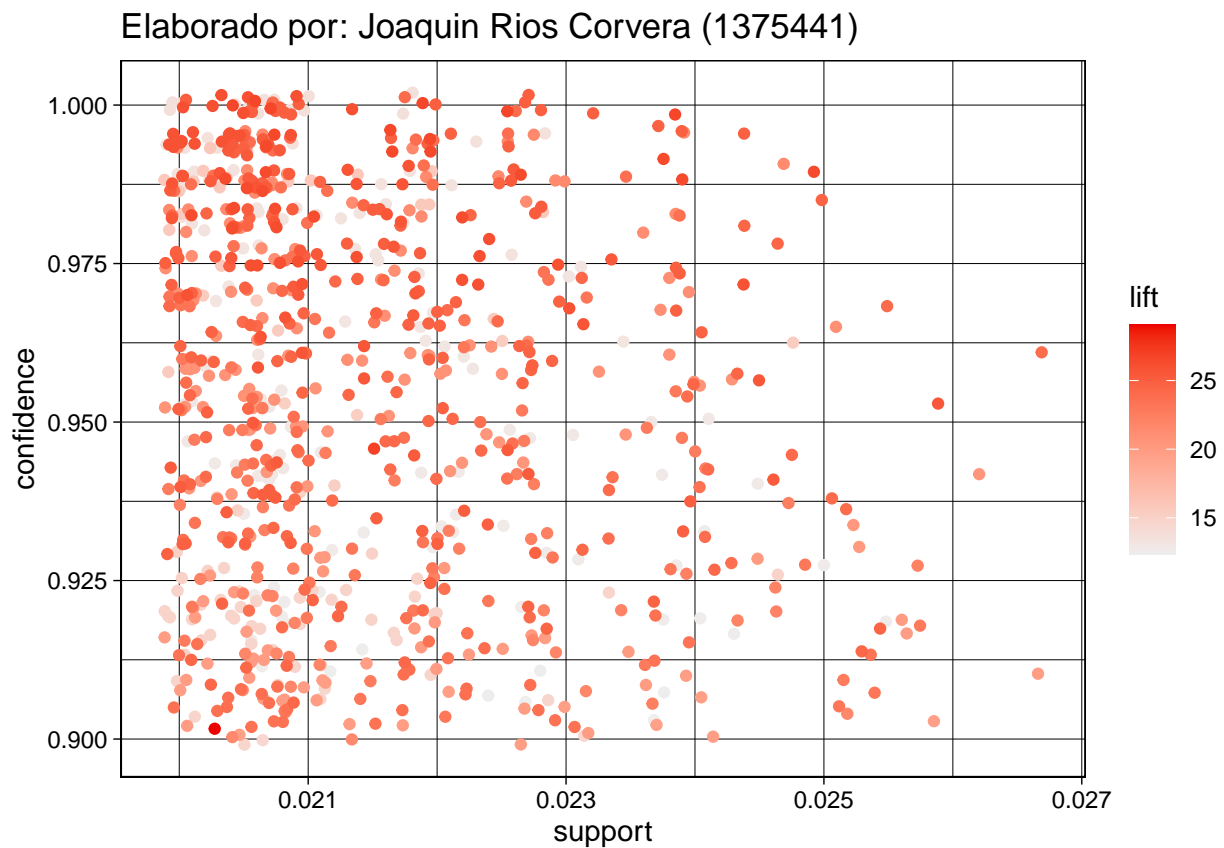
```
summary(rules)
```

```
## set of 883 rules
##
## rule length distribution (lhs + rhs):sizes
##   3   4   5   6   7
## 92 360 311 106  14
##
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.000   4.000   4.000   4.536   5.000   7.000
##
## summary of quality measures:
##      support      confidence      coverage      lift
##   Min.   :0.02000   Min.   :0.9000   Min.   :0.02000   Min.   :12.32
##   1st Qu.:0.02050   1st Qu.:0.9263   1st Qu.:0.02125   1st Qu.:19.65
```

```
## Median :0.02100 Median :0.9552 Median :0.02225 Median :22.68
## Mean :0.02150 Mean :0.9538 Mean :0.02257 Mean :21.02
## 3rd Qu.:0.02225 3rd Qu.:0.9820 3rd Qu.:0.02363 3rd Qu.:24.56
## Max. :0.02662 Max. :1.0000 Max. :0.02925 Max. :29.05
## count
## Min. :160
## 1st Qu.:164
## Median :168
## Mean :172
## 3rd Qu.:178
## Max. :213
##
## mining info:
## data ntransactions support confidence
## mssdtx 8000 0.02 0.9
```

```
try(
  plot(rules, main=paste0("Elaborado por: ", NOMBRE_COMPLETO, " (", DIGITOS_MATRICULA, ")") ))
)
```

To reduce overplotting, jitter is added! Use jitter = 0 to prevent jitter.



Inspecciona las reglas con mayor lift

```
inspect(head(sort(rules, by = "lift")))
```

```
##      lhs                                                    rhs      support con
## [1] {t_e8fa3463-7ef7-4ca5-8e65-2df979012c34,
##      t_eb0063cc-948f-41d4-9dab-bf4b0e70d21e} => {t_4319026b-d2b5-478d-bf37-8f2873645c5c} 0.020375 0
## [2] {t_0e3dec82-10b4-49f1-8c2e-cd19249f7d2c,
##      t_4319026b-d2b5-478d-bf37-8f2873645c5c} => {t_eb0063cc-948f-41d4-9dab-bf4b0e70d21e} 0.021500 0
## [3] {t_312b3509-48c1-4928-a6e5-d3e7a51c216b,
##      t_7533e7ff-568e-448b-904e-fc0c3a9ec87e,
##      t_81b9a202-281e-42a4-b068-e0c0f44e1e4d,
##      t_a0fa9956-ffc3-41c9-a344-e247baa0e321} => {t_1dba18ad-1ddb-4e97-983b-4af57e18d84a} 0.023875 1
## [4] {t_312b3509-48c1-4928-a6e5-d3e7a51c216b,
##      t_7533e7ff-568e-448b-904e-fc0c3a9ec87e,
##      t_7e616ed3-115f-4a97-b90e-d128c392fae0,
##      t_81b9a202-281e-42a4-b068-e0c0f44e1e4d,
##      t_a0fa9956-ffc3-41c9-a344-e247baa0e321} => {t_1dba18ad-1ddb-4e97-983b-4af57e18d84a} 0.020500 1
## [5] {t_312b3509-48c1-4928-a6e5-d3e7a51c216b,
##      t_68888d19-7be6-4cdc-bcfb-be38431f79d7,
##      t_7533e7ff-568e-448b-904e-fc0c3a9ec87e,
##      t_81b9a202-281e-42a4-b068-e0c0f44e1e4d,
##      t_a0fa9956-ffc3-41c9-a344-e247baa0e321} => {t_1dba18ad-1ddb-4e97-983b-4af57e18d84a} 0.021875 1
## [6] {t_312b3509-48c1-4928-a6e5-d3e7a51c216b,
##      t_5718ab08-3a15-4d3f-9e63-42b2f6805e31,
##      t_7533e7ff-568e-448b-904e-fc0c3a9ec87e,
##      t_81b9a202-281e-42a4-b068-e0c0f44e1e4d,
##      t_a0fa9956-ffc3-41c9-a344-e247baa0e321} => {t_1dba18ad-1ddb-4e97-983b-4af57e18d84a} 0.022625 1
```

Graba y publica un video (puede ser un video publicado de forma privada en YouTube u otro servicio similar) de 3 a 5 minutos explicando lo siguiente:

- Nombre y carrera.
- Concepto de association rules.
- Diferencia entre los conceptos de transacciones y reglas.
- Estrategia seguida para descubrir las reglas de interés en el dataset MSSD de este proyecto.
- Interpretación de la primera regla listada en el chunk anterior con el código `inspect(head(sort(rules, by = "lift")))` aclarando qué significan los datos listados en lhs, rhs y la interpretación de todas sus métricas.
- Provee el enlace al video aquí: **NOTE-30** https://youtu.be/dNw9a28z_Fw

Entrega del proyecto final

- Verifica que el notebook corra de principio a fin sin errores y produzca los resultados esperados.
- Verifica que hayas realizado todas las actividades de programación indicadas con **CODE-nn**
- Verifica que hayas respondido a todas las preguntas **NOTE-nn**
- Verifica que hayas producido, publicado y escrito el enlace al video en el notebook.
- Guarda el archivo de markup (rmd) de forma local.
- Genera el PDF del notebook por medio del comando Knit (ya sea directo a PDF o Word y posteriormente Save As/PDF).
- No se recibirán proyectos en formato PDF con texto libre y capturas de pantalla.
- Publica el código de tu notebook a tu repositorio de GitHub por medio de los comandos `commit` y `push`.

- Envía el documento PDF por medio de mensaje personal a @marciano en Slack `naylacommunity` junto con el enlace a tu repositorio de GitHub.

Fecha límite de entrega: Sábado 5 de junio al cierre del día.