

R Notebook - Willy Ardiano - 999017224 - Tarea 3 - Minería de datos

Repositorio de github: <https://github.com/willyrex9912/tarea-3-mineria-de-datos.git>

Instalando dependencias y leyendo archivo

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.4.3
```

```
data <- read_excel("base-de-datos-violencia-intrafamiliar-ano-2024_v3.xlsx")
```

```
#Quitanto y dejando solo las columnas a revisar  
data.frame(1:ncol(data), colnames(data))
```

##	X1.ncol.data.	colnames.data.
## 1	1	HEC_DIA
## 2	2	HEC_MES
## 3	3	HEC_ANO
## 4	4	HEC_DEPTO
## 5	5	HEC_DEPTOMCPIO
## 6	6	HEC_TIPAGRE
## 7	7	NUMERO_BOLETA
## 8	8	DIA_EMISION
## 9	9	MES_EMISION
## 10	10	ANO_EMISION
## 11	11	DEPTO
## 12	12	DEPTO_MCPIO
## 13	13	QUIEN_REPORTA
## 14	14	VIC_SEXO
## 15	15	VIC_EDAD
## 16	16	TOTAL_HIJOS
## 17	17	NUM_HIJ_HOM
## 18	18	NUM_HIJ_MUJ
## 19	19	VIC_ALFAB
## 20	20	VIC_ESCOLARIDAD
## 21	21	VIC_EST_CIV
## 22	22	VIC_GRUPET
## 23	23	VIC_NACIONAL
## 24	24	VIC TRABAJA
## 25	25	VIC_OCUP
## 26	26	VIC_DEDICA
## 27	27	VIC_DISC
## 28	28	TIPO_DISCAQ
## 29	29	VIC_REL_AGR

```

## 30          30          OTRAS_VICTIMAS
## 31          31          VIC_OTRAS_HOM
## 32          32          VIC_OTRAS_MUJ
## 33          33          VIC_OTRAS_N_OS
## 34          34          VIC_OTRAS_N_AS
## 35          35          HEC_AREA
## 36          36          HEC_RECUR_DENUN
## 37          37          INST_DONDE_DENUNCIO
## 38          38          AGR_SEXO
## 39          39          AGR_EDAD
## 40          40          AGR_ALFAB
## 41          41          AGR_ESCOLARIDAD
## 42          42          AGR_EST_CIV
## 43          43          AGR_GURPET
## 44          44          AGR_NACIONAL
## 45          45          AGR_TRABAJA
## 46          46          AGR_OCUP
## 47          47          AGR_DEDICA
## 48          48          AGRESORES_OTROS_TOTAL
## 49          49          AGR_OTROS_HOM
## 50          50          AGR_OTRAS_MUJ
## 51          51          AGR_OTROS_N_OS
## 52          52          AGR_OTRAS_N_AS
## 53          53          INST_DENUN_HECHO
## 54          54          ORGANISMO_JURISDICCIONAL
## 55          55          CONDUCTENTE
## 56          56          LEY_APLICABLE
## 57          57          ARTICULOVIF1
## 58          58          ARTICULOVIF2
## 59          59          ARTICULOVIF3
## 60          60          ARTICULOVIF4
## 61          61          ARTICULOVC1
## 62          62          ARTICULOVC2
## 63          63          ARTICULOVC3
## 64          64          ARTICULOVC4
## 65          65          ARTICULOCODPEN1
## 66          66          ARTICULOCODPEN2
## 67          67          ARTICULOCODPEN3
## 68          68          ARTICULOCODPEN4
## 69          69          ARTICULOTRAS1
## 70          70          ARTICULOTRAS2
## 71          71          ARTICULOTRAS3
## 72          72          ARTICULOTRAS4
## 73          73          MEDIDAS_SEGURIDAD
## 74          74          TIPO_MEDIDA
## 75          75          ORGANISMO_REMITE

```

```
datos <- data[,c('HEC_MES', 'HEC_DEPTO', 'VIC_SEXO', 'VIC_EDAD', 'VIC_ESCOLARIDAD', 'VIC_EST_CIV', 'VIC_
```

Filtrando por departamento de San Marcos

```

datos_dep_12 <- subset(datos, HEC_DEPTO == 12)
# Limpiando el set de datos de San Marcos quitando la columna de departamento
data.frame(1:ncol(datos_dep_12), colnames(datos_dep_12))

```

```
##      X1.ncol.datos_dep_12. colnames.datos_dep_12.
## 1          1          HEC_MES
## 2          2          HEC_DEPTO
## 3          3          VIC_SEXO
## 4          4          VIC_EDAD
## 5          5          VIC_ESCOLARIDAD
## 6          6          VIC_EST_CIV
## 7          7          VIC TRABAJA
## 8          8          VIC_DISC
## 9          9          AGR_SEXO
## 10         10         AGR_EDAD
## 11         11         AGR_ESCOLARIDAD
## 12         12         AGR TRABAJA
## 13         13         AGR_OCUP
## 14         14         AGR_DEDICA
```

```
datos_dep_12 <- datos_dep_12[, -2]
```

Aplicando algoritmo para encontrar reglas

```
library(arules)
```

```
## Warning: package 'arules' was built under R version 4.4.3
```

```
## Cargando paquete requerido: Matrix
```

```
##
```

```
## Adjuntando el paquete: 'arules'
```

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

```
##
```

```
##      abbreviate, write
```

```
reglas_12 <- fim4r(datos_dep_12, method = "fpgrowth", target = "rules", support = 0.2, confidence = 0.5)
```

```
## Warning: Column(s) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 not logical or
```

```
## factor. Applying default discretization (see '? discretizeDF').
```

```
## Warning in discretize(x = c(2, 2, 2, 2, 1, 2, 2, 2, 2, 1, 2, 2, 1, 1, 1, : The calculated breaks are
```

```
##      Only unique breaks are used reducing the number of intervals. Look at ? discretize for details.
```

```
## Warning in discretize(x = c(1, 9, 2, 9, 9, 2, 9, 9, 9, 9, 9, 2, 2, 1, 2, : The calculated breaks are
```

```
##      Only unique breaks are used reducing the number of intervals. Look at ? discretize for details.
```

```
## Warning in discretize(x = c(2, 2, 2, 2, 1, 2, 2, 2, 2, 1, 2, 2, 1, 1, 1, : The calculated breaks are
```

```
##      Only unique breaks are used reducing the number of intervals. Look at ? discretize for details.
```

```
## Warning in discretize(x = c(2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, : The calculated breaks are
```

```
##      Only unique breaks are used reducing the number of intervals. Look at ? discretize for details.
```

```
## Warning in discretize(x = c(2, 2, 1, 1, 2, 1, 1, 1, 1, 2, 1, 1, 2, 1, 2, : The calculated breaks are
##   Only unique breaks are used reducing the number of intervals. Look at ? discretize for details.

## Warning in discretize(x = c(2, 2, 2, 1, 2, 2, 2, 1, 1, 2, 1, 2, 2, 1, 2, : The calculated breaks are
##   Only unique breaks are used reducing the number of intervals. Look at ? discretize for details.

## Warning in discretize(x = c(3, 3, 3, NA, 1, 9, 3, NA, NA, 1, NA, 9, 3, NA, : The calculated breaks are
##   Only unique breaks are used reducing the number of intervals. Look at ? discretize for details.

## fim4r.fpgrowth
##
## Parameter specification:
##  supp conf target report
##    20   50  rules    scl
##
## Data size: 2113 transactions and 28 items
## Result: 3856 rules
```

```
reglas_12_df <- as(reglas_12, Class = 'data.frame')
```

Patron 1: El primer patrón que puedo observar al filtrar los datos por el departamento 12 (San Marcos) es “VIC_DISC=[2,9]” Parece ser que la mayor parte de víctimas no tienen una discapacidad. Por un lado es bueno porque significa que las agresiones no tienen nada que ver con la discapacidad de alguien, en caso contrario sería información muy trágica. Si bien es cierto que no parece haber relevancia con las personas que tienen discapacidad, tampoco significa que no exista. Queda abierta la posibilidad de que alguna persona con discapacidad reciba alguna agresión.

Patron 2: El segundo patron que puedo observar es “VIC_EST_CIV=[2,9]”. Por el diccionario de datos podemos deducir que la mayor parte de víctimas que sufren algún tipo de agresión son personas que se han casado o unido con otra persona. Eso es interesante porque nos dice que los solteros en gran gran porcentaje no sufren de algún tipo de agresión, al menos no en los registros que se han registrado.

Patron 3: El tercer patrón que puedo observar es “VIC_TRABAJA=[2,9],AGR_TRABAJA=[1,9]”. Al parecer las víctimas son personas que no trabajan por un salario o ingreso mientras que los agresores son personas que si trabajan por un salario. Esto parece lógico al pensar que muchas veces los problemas familiares surgen debido a problemas económicas o discusiones sobre temas económicos que se puedan dar. Por lo que estadísticamente se puede fundamentar que una persona tiene más riesgo de sufrir algún tipo de agresión si no cuenta con un trabajo que le permita obtener un sueldo o salario.

Patron 4: Me parece muy muy interesante el patron “VIC_SEXO=[1,2]”. Esperaba encontrar una tendencia de violencia contra la mujer, pero es interesante que al observar la mayor parte de reglas y patrones indican que las agresiones se dan tanto en hombres como en mujeres. De hecho me surgió la duda sobre si es cierto y revisé los registros, y es interesante porque en efecto los registros de hombres y mujeres están bastante nivelados. Confirma los calculos arrojados por el algoritmo al menos en los datos del departamento de San Marcos.