HackForGood - El síndrome de BurnOut (Primera limpieza)

MAD09 - Maslash

2024-03-15

PRIMERA LIMPIEZA DE LOS DATOS

En esta primera limpieza lo que haremos será adaptar los resultados obtenidos a través de la encuesta para poder analizarlos más fácilmente.

```
# Leemos el fichero de datos, que está en formato csv
encuesta <- read.csv('Síndrome de Burnout.csv')
head(encuesta)
```

```
##                Marca.temporal X.Eres.estudiante.de.42.
## 1 2024/03/15 2:10:23 p. m. CET                      No
## 2 2024/03/15 2:11:03 p. m. CET                      No
## 3 2024/03/15 2:11:23 p. m. CET                      No
## 4 2024/03/15 2:11:55 p. m. CET                      Si
## 5 2024/03/15 2:12:19 p. m. CET                      Si
## 6 2024/03/15 2:12:20 p. m. CET                      Si
##   X.Te.sientes.motivado.en.tus.estudios.proyectos.
## 1                                               No
## 2                                   Si, mucho
## 3                                               No
## 4                                   Si, mucho
## 5                                 Sí, pero poco
## 6                                 Sí, pero poco
##   X.Elegiste.tus.estudios.actuales. X.Te.gusta.lo.que.estudias.
## 1            Si, era lo que quería                          No
## 2            Si, era lo que quería                          Si
## 3            Si, era lo que quería                          Si
## 4            Si, era lo que quería                          Si
## 5            Si, era lo que quería                          Si
## 6              No, acabe de rebote                          Si
##   X.Sientes.que.aprovechas.el.tiempo. X.Eres.una.persona.organizada.
## 1                                  Si                     Si, mucho
## 2                                  Si                     Si, mucho
## 3                                  No                      No, nada
## 4                                  Si                   Si, un poco
## 5                                  Si                   Si, un poco
## 6                                  No                      No, nada
##   X.Te.sientes.cómodo.con.tus.compañeros.en.el.ambiente.de.estudio.
## 1                                                                Si
## 2                                                                Si
```

```
## 3                                                                 Si
## 4                                                                 Si
## 5                                                                 Si
## 6                                                                 Si
##   X.Cuántas.horas.duermes.al.día...en.horas.
## 1                                         6
## 2                                         3
## 3                                         7
## 4                                         6
## 5                                         8
## 6                                         7
##   X.Sientes.que.duermes.lo.suficiente. X.Cuántas.comidas.haces.diariamente.
## 1                                   No                             Mas de 3
## 2                                   No                            Menos de 3
## 3                                   Si                             Mas de 3
## 4                                   Si                            Menos de 3
## 5                                   Si                                     3
## 6                                   Si                                     3
##   X.Cuántas.bebidas.energéticas.o.café.consumes.diariamente.
## 1                                                      Nada
## 2                                                         1
## 3                                                         1
## 4                                                      Nada
## 5                                                         1
## 6                                                      Nada
##   X.Cada.cuánto.realizas.actividad.física.meditación.semanalmente.
## 1                                                        5 o mas
## 2                                                     1 - 2 dias
## 3                                                     3 - 4 dias
## 4                                                     1 - 2 dias
## 5                                                     1 - 2 dias
## 6                                                     3 - 4 dias
##   X.Cuánto.tiempo.dedicas.a.los.estudios.diariamente...en.horas.
## 1                                                              6
## 2                                                              4
## 3                                                              2
## 4                                                              8
## 5                                Contando el tiempo en telefonica 8
## 6                                                              6
##   X.Te.sientes.a.gusto.con.la.metodología.de.estudio.que.tiene.tu.sistema.educativo.
## 1                                                                                 No
## 2                                                                                 Si
## 3                                                                                 No
## 4                                                                                 Si
## 5                                                                                 Si
## 6                                                                                 Si
##   X.Te.sientes.acompañado.en.el.estudio.
## 1                                     No
## 2                                     Si
## 3                                     No
## 4                                     Si
## 5                                     Si
## 6                                     No
##   X.Sientes.que.estás.en.un.ambiente.competitivo.
```

```
## 1                                              No
## 2                                              Si
## 3                                              Si
## 4                                              No
## 5                                              No
## 6                                              Si
##   X.Te.sientes.cómodo.en.un.ambiente.competitivo.
## 1                                              Si
## 2                                              Si
## 3                                              Si
## 4                                              No
## 5                                              Si
## 6                                              Si
##   X.Te.ves.capaz.de.realizar.trabajos.de.forma.autónoma.
## 1                                                     Si
## 2                                                     Si
## 3                                                     Si
## 4                                                     Si
## 5                                                     Si
## 6                                                     Si
##   X.Sientes.presión.cuando.se.te.acumulan.las.tareas.
## 1                                                  No
## 2                                                  Si
## 3                                                  No
## 4                                                  Si
## 5                                                  No
## 6                                                  Si
##   X.Crees.que.tienes.el.síndrome.de.burnout.
## 1                                         Si
## 2                                         No
## 3                                         No
## 4                                         Si
## 5                                         No
## 6                                         Si
```

Quitamos la marca temporal, pues no aporta ninguna información de utilidad al estudio.

```
datos_encuesta <- encuesta[,-c(1)]
head(datos_encuesta)
```

```
##   X.Eres.estudiante.de.42. X.Te.sientes.motivado.en.tus.estudios.proyectos.
## 1                       No                                              No
## 2                       No                                       Si, mucho
## 3                       No                                              No
## 4                       Si                                       Si, mucho
## 5                       Si                                  Sí, pero poco
## 6                       Si                                  Sí, pero poco
##   X.Elegiste.tus.estudios.actuales. X.Te.gusta.lo.que.estudias.
## 1            Si, era lo que quería                           No
## 2            Si, era lo que quería                           Si
## 3            Si, era lo que quería                           Si
## 4            Si, era lo que quería                           Si
## 5            Si, era lo que quería                           Si
```

```
## 6               No, acabe de rebote                           Si
##   X.Sientes.que.aprovechas.el.tiempo. X.Eres.una.persona.organizada.
## 1                                  Si                      Si, mucho
## 2                                  Si                      Si, mucho
## 3                                  No                       No, nada
## 4                                  Si                    Si, un poco
## 5                                  Si                    Si, un poco
## 6                                  No                       No, nada
##   X.Te.sientes.cómodo.con.tus.compañeros.en.el.ambiente.de.estudio.
## 1                                                                Si
## 2                                                                Si
## 3                                                                Si
## 4                                                                Si
## 5                                                                Si
## 6                                                                Si
##   X.Cuántas.horas.duermes.al.día...en.horas.
## 1                                          6
## 2                                          3
## 3                                          7
## 4                                          6
## 5                                          8
## 6                                          7
##   X.Sientes.que.duermes.lo.suficiente. X.Cuántas.comidas.haces.diariamente.
## 1                                   No                           Mas de 3
## 2                                   No                          Menos de 3
## 3                                   Si                           Mas de 3
## 4                                   Si                          Menos de 3
## 5                                   Si                                   3
## 6                                   Si                                   3
##   X.Cuántas.bebidas.energéticas.o.café.consumes.diariamente.
## 1                                                       Nada
## 2                                                          1
## 3                                                          1
## 4                                                       Nada
## 5                                                          1
## 6                                                       Nada
##   X.Cada.cuánto.realizas.actividad.física.meditación.semanalmente.
## 1                                                        5 o mas
## 2                                                      1 - 2 dias
## 3                                                      3 - 4 dias
## 4                                                      1 - 2 dias
## 5                                                      1 - 2 dias
## 6                                                      3 - 4 dias
##   X.Cuánto.tiempo.dedicas.a.los.estudios.diariamente...en.horas.
## 1                                                              6
## 2                                                              4
## 3                                                              2
## 4                                                              8
## 5                                Contando el tiempo en telefonica 8
## 6                                                              6
##   X.Te.sientes.a.gusto.con.la.metodología.de.estudio.que.tiene.tu.sistema.educativo.
## 1                                                                                 No
## 2                                                                                 Si
## 3                                                                                 No
```

```
## 4                                                                                 Si
## 5                                                                                 Si
## 6                                                                                 Si
##   X.Te.sientes.acompañado.en.el.estudio.
## 1                                    No
## 2                                    Si
## 3                                    No
## 4                                    Si
## 5                                    Si
## 6                                    No
##   X.Sientes.que.estás.en.un.ambiente.competitivo.
## 1                                             No
## 2                                             Si
## 3                                             Si
## 4                                             No
## 5                                             No
## 6                                             Si
##   X.Te.sientes.cómodo.en.un.ambiente.competitivo.
## 1                                             Si
## 2                                             Si
## 3                                             Si
## 4                                             No
## 5                                             Si
## 6                                             Si
##   X.Te.ves.capaz.de.realizar.trabajos.de.forma.autónoma.
## 1                                                    Si
## 2                                                    Si
## 3                                                    Si
## 4                                                    Si
## 5                                                    Si
## 6                                                    Si
##   X.Sientes.presión.cuando.se.te.acumulan.las.tareas.
## 1                                                 No
## 2                                                 Si
## 3                                                 No
## 4                                                 Si
## 5                                                 No
## 6                                                 Si
##   X.Crees.que.tienes.el.síndrome.de.burnout.
## 1                                        Si
## 2                                        No
## 3                                        No
## 4                                        Si
## 5                                        No
## 6                                        Si
```

Al haber diseñado la encuensta de forma que todas las preguntas fuesen de respuesta obligatoria, no será necesario comprobar si hay NAs.

El nombre de las columnas es demasiado largo, pues al ser una encuesta las variables son las preguntas que se han realizado, por lo que sería conveniente cambiar los nombres de la mayoria de estás variables a unos más cortos, para que su manipulación sea más cómoda.

```r
names(datos_encuesta) = c('Estudiante.42', 'Motivacion', 'Eleccion.estudios',
                          'Gustar.estudios', 'Aprovechar.tiempo', 'Organizacion',
                          'Comodo.compañeros', 'Horas.sueño', 'Dormir.suficiente',
                          'Comidias.diarias', 'Bebidas.energéticas',
                          'Actividad.fisica', 'Tiempo.estudios', 'Metodología',
                          'Acompañado', 'Ambiente.competitivo',
                          'Comodo.competicion', 'Autonomía', 'Presion.tareas',
                          'Burnout')
head(datos_encuesta)
```

```
##   Estudiante.42    Motivacion     Eleccion.estudios Gustar.estudios
## 1            No            No Si, era lo que quería              No
## 2            No     Si, mucho Si, era lo que quería              Si
## 3            No            No Si, era lo que quería              Si
## 4            Si     Si, mucho Si, era lo que quería              Si
## 5            Si Sí, pero poco Si, era lo que quería              Si
## 6            Si Sí, pero poco   No, acabe de rebote              Si
##   Aprovechar.tiempo Organizacion Comodo.compañeros Horas.sueño
## 1                Si    Si, mucho                Si           6
## 2                Si    Si, mucho                Si           3
## 3                No     No, nada                Si           7
## 4                Si  Si, un poco                Si           6
## 5                Si  Si, un poco                Si           8
## 6                No     No, nada                Si           7
##   Dormir.suficiente Comidias.diarias Bebidas.energéticas Actividad.fisica
## 1                No         Mas de 3                Nada          5 o mas
## 2                No       Menos de 3                   1       1 - 2 dias
## 3                Si         Mas de 3                   1       3 - 4 dias
## 4                Si       Menos de 3                Nada       1 - 2 dias
## 5                Si                3                   1       1 - 2 dias
## 6                Si                3                Nada       3 - 4 dias
##                     Tiempo.estudios Metodología Acompañado
## 1                                 6          No         No
## 2                                 4          Si         Si
## 3                                 2          No         No
## 4                                 8          Si         Si
## 5 Contando el tiempo en telefonica 8          Si         Si
## 6                                 6          Si         No
##   Ambiente.competitivo Comodo.competicion Autonomía Presion.tareas Burnout
## 1                   No                Si        Si             No      Si
## 2                   Si                Si        Si             Si      No
## 3                   Si                Si        Si             No      No
## 4                   No                No        Si             Si      Si
## 5                   No                Si        Si             No      No
## 6                   Si                Si        Si             Si      Si
```

Convertimos todas las variables a categoricas.

```r
datos_encuesta2 <- as.data.frame(lapply(datos_encuesta, as.factor))
summary(datos_encuesta2)
```

```
##  Estudiante.42         Motivacion             Eleccion.estudios Gustar.estudios
```

```
##  No:45         No           : 6   No, acabe de rebote  : 9      No: 4
##  Si:57         Si, mucho    :71   Si, era lo que quería:93      Si:98
##                Sí, pero poco:25
##
##
##
##
##  Aprovechar.tiempo      Organizacion Comodo.compañeros  Horas.sueño
##  No:33             No, nada   :26    No: 8             7      :27
##  Si:69             Si, mucho  :26    Si:94             6      :20
##                    Si, un poco:50                      8      :11
##                                                        4      : 5
##                                                        9      : 5
##                                                        7-8    : 4
##                                                        (Other):30
##  Dormir.suficiente   Comidias.diarias Bebidas.energéticas   Actividad.fisica
##  No:48             3         :47      1      :36          1 - 2 dias:34
##  Si:54             Mas de 3  :31      2      :21          3 - 4 dias:25
##                    Menos de 3:24      3 o más: 9          5 o mas   :24
##                                       Nada   :36          Nada      :19
##
##
##
##  Tiempo.estudios Metodología Acompañado Ambiente.competitivo Comodo.competicion
##  3      :11      No:27       No:27      No:41                No:34
##  4      :11      Si:75       Si:75      Si:61                Si:68
##  5      :11
##  6      : 9
##  8      : 8
##  2      : 7
##  (Other):45
##  Autonomía Presion.tareas Burnout
##  No:14     No:29          No:72
##  Si:88     Si:73          Si:30
##
##
##
##
##
```

Vamos a cambiar las categorías de la variable 'Eleccion.estudios' a simplemente si o no, porque la otra parte era explicativa para que la gente lo entendiese mejor a la hora de contestar.

```
library(dplyr)
```

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

## The following objects are masked from 'package:stats':
##
##     filter, lag

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

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

```
datos_encuesta2 <- datos_encuesta2 %>%
  mutate(Eleccion.estudios = recode(Eleccion.estudios, "No, acabe de rebote" = "No",
                                    "Si, era lo que quería" = "Si"))
```

Las variables que requieren de una limpieza más profunda son aquellas cuya respuesta es abierta. Para ver cuántas categorías hay y si hay algunas que representen lo mismo pero recogido de diferente forma realizamos un conteo de las categorías que hay en dicha variable.

Empezamos con la variable 'Horas.sueño'

```
# Obtener conteos de los valores presentes en la variable Horas.sueño en el dataframe
conteo_Horas.sueño <- table(datos_encuesta2$Horas.sueño)

# Mostrar el resultado
print(conteo_Horas.sueño)
```

```
##
##              1             10          10-6       10 horas             3
##              1              1             1             1             2
##            3-5       3h 26min             4   4 - 7 4 horas aprox.
##              1              1             5             1             1
##              5            5-6             6          6-7h          6-8
##              3              2            20             1             1
##          6 o 7          6 ó 7           6,5          6.33            6h
##              1              1             2             1             1
##              7            7-8       7 horas           7,5         7,5h
##             27              4             1             1             1
##            7.5             7h             8             8       8 horas
##              1              1            11             1             1
##              9
##              5
```

Como se puede observar, hay muchas categorías que representan intervalos, horas que no son enteras, o incluso, la presencia de palabras como 'horas', todo esto no nos interesa a la hora de estudio. Para solucionar este problema seguiremos las siguientes pautas: - Solo dejaremos números enteros; nada de palabras, letras, espacios o símbolos como '-' - Los número decimales los truncaremos -> Por ejemplo, si tenemos 6.5 nos quedamos con el 6 - Los intervalos se enfrentarán de dos formas: - Calcularemos la media del intervalo, si es un número entero nos quedamos con él. - Si no es un número entero; truncamos

```
# Recodificar las categorías
datos_encuesta2 <- datos_encuesta2 %>%
  mutate(Horas.sueño = recode(Horas.sueño, "10" = "10", "10 horas" = "10"))

datos_encuesta2 <- datos_encuesta2 %>%
  mutate(Horas.sueño = recode(Horas.sueño, "4" = "4", "4 horas aprox." = "4",
                              "3-5" = "4"))

datos_encuesta2 <- datos_encuesta2 %>%
  mutate(Horas.sueño = recode(Horas.sueño, "6" = "6", "6h" = "6", "6-7h" = "6",
                              "6 o 7" = "6", "6 ó 7" = "6", "6,5" = "6",
```

```
                              "6.33" = "6"))

datos_encuesta2 <- datos_encuesta2 %>%
  mutate(Horas.sueño = recode(Horas.sueño, "7" = "7", "7 horas" = "7",
                              "7h" = "7", "6-8" = "7", "7-8" = "7",
                              "7,5" = "7", "7,5h" = "7", "7.5" = "7"))

datos_encuesta2 <- datos_encuesta2 %>%
  mutate(Horas.sueño = recode(Horas.sueño, "8" = "8", "8 " = "8",
                              "8 horas" = "8", "10-6" = "8"))

datos_encuesta2 <- datos_encuesta2 %>%
  mutate(Horas.sueño = recode(Horas.sueño, "3h 26min" = "3"))

datos_encuesta2 <- datos_encuesta2 %>%
  mutate(Horas.sueño = recode(Horas.sueño, "4 - 7" = "5", "5-6" = "5"))

# Volvemos a obtener conteos de los valores presentes en la variable para
# comprobar que los cambios se han realizado bien
conteo1_Horas.sueño <- table(datos_encuesta2$Horas.sueño)

# Mostrar el resultado
print(conteo1_Horas.sueño)
```

```
##
##  1 10  8  3  4  5  6  7  9
##  1  2 14  3  7  6 27 37  5
```

Continuamos con la variable 'Tiempo.estudios'

```
conteo_Tiempo.estudios <- table(datos_encuesta2$Tiempo.estudios)

print(conteo_Tiempo.estudios)
```

```
##
##                                          0
##                                          4
##                                        0.2
##                                          1
##                                        0.5
##                                          1
##                                          1
##                                          4
##                                        1-2
##                                          2
##          1 hora o más dependiendo de la semana
##                                          1
##                                      1 o 2
##                                          1
##                                         10
##                                          1
##                                         12
```

```
##                                             2
##                                             2
##                                             7
## 2-3 (sin contar el tiempo que estamos en clase)
##                                             1
##                                       20 horas
##                                             1
##                                          25 h
##                                             1
##                                        2horas
##                                             1
##                                             3
##                                            11
##                                             4
##                                            11
##                                           4-5
##                                             1
##                                         4 - 8
##                                             1
##                                       4 horas
##                                             1
##                                            40
##                                             1
##                                             5
##                                            11
##                                   5 - 6 horas
##                                             1
##                                             6
##                                             9
##                                          6-10
##                                             1
##                                           6-7
##                                             1
##                                          6-7h
##                                             2
##                                            6h
##                                             1
##                                         6h-10h
##                                             1
##                                             7
##                                             4
##                                            7h
##                                             1
##                                             8
##                                             8
##                                       8 horas
##                                             1
##                                             9
##                                             2
##                                          9-10
##                                             1
##              Contando el tiempo en telefonica 8
##                                             1
##                                 Entre 8 y 12
```

```
##                                         1
##                                   No lo sé
##                                         1
##                   Unas 8 seis días a la semana
##                                         1
```

La forma de limpieza de esta variables es exactamente igual que la anterior con el caso anterior excepto en un caso puntual. Cuando el tiempo de estudio es 0.x lo asociamos a la categoría del 1 en vez de a la de 0, pues no es lo mismo estudiar menos de una hora que no estudiar absolutamente nada.

```
# Recodificar las categorías
datos_encuesta2 <- datos_encuesta2 %>%
  mutate(Tiempo.estudios = recode(Tiempo.estudios, "8" = "8", "8 horas" = "8",
                                  "Contando el tiempo en telefonica 8" = "8",
                                  "Unas 8 seis días a la semana" = "8",
                                  "6h-10h" = "8", "6-10" = "8"))

datos_encuesta2 <- datos_encuesta2 %>%
  mutate(Tiempo.estudios = recode(Tiempo.estudios, "0.2" = "1", "0.5" = "1",
                                  "1-2" = "1", "1 o 2" = "1",
                                  "1 hora o más dependiendo de la semana" = "1"))

datos_encuesta2 <- datos_encuesta2 %>%
  mutate(Tiempo.estudios = recode(Tiempo.estudios, "4-5" = "4", "4 horas" = "4"))

datos_encuesta2 <- datos_encuesta2 %>%
  mutate(Tiempo.estudios = recode(Tiempo.estudios, "2horas" = "2",
                                  "2-3 (sin contar el tiempo que estamos en clase)" = "2"))

datos_encuesta2 <- datos_encuesta2 %>%
  mutate(Tiempo.estudios = recode(Tiempo.estudios, "5 - 6 horas" = "5"))

datos_encuesta2 <- datos_encuesta2 %>%
  mutate(Tiempo.estudios = recode(Tiempo.estudios, "6-7" = "6", "6-7h" = "6",
                                  "6h" = "6", "4 - 8" = "6"))

datos_encuesta2 <- datos_encuesta2 %>%
  mutate(Tiempo.estudios = recode(Tiempo.estudios, "7h" = "7"))

datos_encuesta2 <- datos_encuesta2 %>%
  mutate(Tiempo.estudios = recode(Tiempo.estudios, "9-10" = "9"))

datos_encuesta2 <- datos_encuesta2 %>%
  mutate(Tiempo.estudios = recode(Tiempo.estudios, "Entre 8 y 12" = "10"))
```

```
conteo1_Tiempo.estudios <- table(datos_encuesta2$Tiempo.estudios)

print(conteo1_Tiempo.estudios)
```

```
##
##        0        1       10       12        2 20 horas     25 h        3
##        4       10        2        2        9        1        1       11
##        4        6       40        5        8        7        9 No lo sé
##       13       14        1       12       13        5        3        1
```

Gracias a este conteo podemos observar que hay 4 variables cuyos datos de esta variable no se ajustan a los parámetros, pues la pregunta es cuántas horas estudia una persona diariamente y rangos como 20, 25, 40 o No lo sé, no sirven para darle generalidad.

Antes de eliminar estas observaciones, observemos cuántas tenemos en total.

```
# Usando nrow()
num_observaciones <- nrow(datos_encuesta2)
print(num_observaciones)
```

```
## [1] 102
```

Eliminamos las observaciones que hemos detectado como 'defectuosas' y volvemos a obtener el número total de observaciones después de este cambio para comprobar si efectivamente se han eliminado 4 y solo 4 filas.

```
datos_encuesta3 <- datos_encuesta2[!(datos_encuesta2$Tiempo.estudios == "25 h" |
                                     datos_encuesta2$Tiempo.estudios == "20 horas" |
                                     datos_encuesta2$Tiempo.estudios == "40" |
                                     datos_encuesta2$Tiempo.estudios == "No lo sé"), ]


num_observaciones2 <- nrow(datos_encuesta3)
print(num_observaciones2)
```

```
## [1] 98
```

Ya vemos que efectivamente se han eliminado solo 4 filas, veamos si justo son las que queriamos volviendo a hacer un conteo de los datos.

```
# Obtener conteos de los valores presentes en la variable Dias.VM en el data frame
conteo2_Tiempo.estudios <- table(datos_encuesta3$Tiempo.estudios)

# Mostrar el resultado
print(conteo2_Tiempo.estudios)
```

```
##
##        0        1       10       12        2 20 horas     25 h        3
##        4       10        2        2        9        0        0       11
##        4        6       40        5        8        7        9 No lo sé
##       13       14        0       12       13        5        3        0
```

Descargamos en forma de excel la base de datos ya limpiada para poder realizar el análisis de datos y clusterizar.

```
# Suponiendo que 'data' es tu DataFrame modificado
library(writexl)
```

```
## Warning: package 'writexl' was built under R version 4.3.3
```

```
write_xlsx(datos_encuesta3, "archivo_modificado0.xlsx")
```